The genomic epidemiology of nontyphoidal *Salmonella* serovars and their association with invasive disease

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Caisey Victoria Pulford

August 2020

Abstract

Title of thesis: The genomic epidemiology of nontyphoidal *Salmonella* serovars and their association with invasive disease

Author: Caisey Victoria Pulford

Nontyphoidal *Salmonella* (NTS) is a globally important bacterial pathogen, typically associated with foodborne gastroenteritis. However, NTS can also invade normally sterile sites in humans, resulting in an extraintestinal infection known as invasive nontyphoidal *Salmonella* (iNTS) disease. The majority of cases of iNTS disease in humans are caused by a small minority of *Salmonella* pathovariants, which exhibit genomic signatures of adaptation to a systemic lifestyle within a restricted host-range. With the increasing availability of whole genome sequencing (WGS)-based pathogen surveillance, it has become possible to use genomic approaches to investigate the adaptive process that occurred when NTS serovars evolved to exploit a new niche. **I aimed to determine the temporal, geographical and evolutionary trends in nontyphoidal *Salmonella* serovars, in order to understand the genomic signatures acquired on the trajectory towards extraintestinal adaptation.**

My genome-based analysis began with a phylogenetic approach to differentiate a broad range of NTS serovars. Specifically, I assembled a unique dataset of 133 reptilian isolates to study *S. enterica* species-level evolution and ecology, with a focus on *Salmonella* that colonised the guts of venomous snakes. The serovars belonged to four of the six *S. enterica* subspecies and included a small number of multidrug resistant isolates of clinical relevance. The subspecies *enterica* isolates were distributed between two distinct phylogenetic clusters, clade A and clade B, and exhibited different levels of metabolic flexibility reflecting distinctive host-ranges. My species-level analysis provided a context for a more detailed study of two epidemiologically important serovars responsible for iNTS disease in humans; *S.* Typhimurium and *S.* Panama.

By studying the stepwise evolution of 680 *S.* Typhimurium isolates that caused bloodstream infection in sub-Saharan Africa, I uncovered the timeline of crucial loss-of-function genetic events that facilitated the emergence of the invasive ST313 *S.* Typhimurium pathovariant in the immunocompromised human niche. The findings link the acquisition of particular nucleotide changes with gene function during the evolution of ST313 lineage 2 (ST313 L2). A recently emerged pan-susceptible ST313 lineage (ST313 L3) was discovered in Malawi, with evidence of genome degradation in important virulence genes which had not occurred in other ST313 lineages. My work describes the evolutionary dynamics of *S*. Typhimurium ST313 and highlights epidemiological shifts in the circulation of ST313 lineages in Malawi.

To compare and contrast with my work in *S.* Typhimurium, I explored an understudied serovar responsible for extraintestinal infection globally. Specifically, I described the population structure, phylodynamics and invasiveness potential of 489 *S.* Panama isolates from 27 countries. The findings revealed 6 geographically associated clades, and regional trends in AMR. Finally, I determined the genetic markers of invasiveness that varied between *S.* Panama clades, highlighting the importance of prophage presence and genome degradation, as was the case in *S.* Typhimurium.

Taken together, my study provides new insights into the evolution of niche-adaptation in a globally important pathogenic genus. My findings highlight the need for an integrated approach for pathogen surveillance and scientific research that combines the power of genome-level discrimination, with epidemiological rigour, clinical observation and functional analysis.

Acknowledgements

Firstly, I would like to thank my primary supervisor, Jay Hinton for his continuous expertise and advice. Jay sparked my interest in infectious disease epidemics during my undergraduate years when he first welcomed me into his lab group. Since then, he has never faltered to provide endless opportunities which have given me an incredible foundation for my scientific career. I would additionally like to thank my secondary supervisor, Kate Baker, who has been an outstanding mentor and role model over the past four years. I will be forever grateful to Kate for her endless guidance, support and inspiring discussions. I would also like to reserve a special thanks to Kate for pushing me to find my scientific voice and always believing that I could become an empowered and confident woman in science.

I am fortunate to have been part of a stimulating and friendly academic environment in Lab H and I would like to thank every former and past member for making the past 4 years so fun! Specifically, I thank the Hinton group and the Bakery, who have always found time to listen to my ideas and offer their expertise including; Blanca Perez-Sepulveda, Nicolas Wenner, Rocío Canals, Will Rowe, Alex Predeus, Wai Yee Fong, Xiaojun Zhu, Yan Li, Ella Rodwell, Hermione Webster, Arthur Bowers-Barnard, Martha Redway, Jessica Bevington, Hannah Trivett and Tanya Horne (Hinton Group); and Rebecca Bengtsson, P.Malaka De Silva, Rebecca Bennett and George Stenhouse (The Bakery). I am particularly indebted to Blanca, for her advice, friendship, reassurance and motivation over the past four years. On a similar note, I am thankful to Ella, for her ability to listen as a scientist and as a friend and for helping me untangle my thoughts; to Becky Bennett, for encouraging me to trust my ideas even after they got us marooned on a beach; and to Hermione, for always knowing when to check in and for getting excited about every single result. I am grateful to Nico, for teaching me how to navigate a research lab and for all of the *very good* and inspiring conversations; to Will, who built the foundations of my bioinformatics knowledge and encouraged my endless computational questions; to Rebecca Bengtsson, for taking time out of her own research to help me troubleshoot and overcome writers block; to Rocío for gifting me with some of her infinite *Salmonella* knowledge and for taking the time to listen; to Lizeth, for her good humoured life advice and to Charlotte Chong, for knowing when I needed a chat and a sprite.

I have been blessed to have been part of multiple globally collaborative projects. I thank every person who helped to make these collaborations so fruitful and enjoyable. I am grateful to MLW particularly Nick Feasey, Melita Gordon and Benjamin Kumwenda for sharing so much of their knowledge and welcoming me so warmly in Malawi; to LSTM particularly Rob Harrison and Paul Rowley for sharing their expertise in venomous snakes; to the Institut Pasteur particularly Francois-Xavier Weill for his infinite wisdom and Alicia Tran Dien and Elise Hansen for showing me Paris; to Public Health England particularly Marie Anne Chattaway, Tim Dallman, Claire Jenkins and Lesley Larkin for taking me under their wing and encouraging my love for public health; and to the Doherty Institute particularly Deborah Williamson and Danielle Ingle for helping this project reach all corners of the globe.

Finally, I would like to thank my incredible family, partner and friends, who have cheered me on to the finish line. I am eternally thankful to my wonderful Mum, without whom none of this could have been possible. Her unconditional love, understanding and belief in me has encouraged me every step of the way; to my amazing Dad who always taught me to have confidence and be strong, and who has given me an endless supply of encouragement and vegetables; to Mike for teaching me to explore the world from every perspective; to my wonderful partner Karl, who I met in the middle of my PhD journey and who has provided love, patience, understanding and humour ever since; and to my lifelong friends, particularly Anna, Rachel and Katherine who have kept me sane, and waited ever so patiently for this thesis to be complete, so that we can finally open the bubbles!

This thesis is dedicated to by beautiful Nan,

Annie Elizabeth Byrne,

who believed that I could achieve my dreams.

"Age Quod Agis"

Publications, presentations and awards

Below is a list of publications, presentations and awards given throughout the course of this PhD. Publications where I am a first or joint-first author are indicated with an asterisk.

Publications in peer-reviewed journals

Pulford CV*, Wenner N, Redway ML, Rodwell EV, Webster HJ, Escudero R, Kröger C, Canals R, Rowe W, Lopez J, Hall N, Rowley PD, Timofte D, Harrison RA, Baker, KS, Hinton JCD. (2019). The diversity, evolution and ecology of *Salmonella* in venomous snakes. PLOS Neglected Tropical Diseases 13(6): e0007169.

Pulford CV*, Perez-Sepulveda BM, Rodwell EV, Weill FX, Baker KS, Hinton JCD. (2019). *Salmonella enterica* Serovar Panama, an Understudied Serovar Responsible for Extraintestinal Salmonellosis Worldwide. Infection and Immunity 87 (9) e00273-19.

Wilson CN*, **Pulford CV***, Akoko J, Perez Sepulveda B, Predeus AV, Bevington J, Duncan P, Hall N, Wigley P, Feasey NA, Pinchbeck G, Hinton JCD, Gordon MA, Fèvre EM. (2020). *Salmonella* identified in pigs in Kenya and Malawi reveals the potential for zoonotic transmission of *Salmonella* in emerging pork markets. PLOS Neglected Tropical Diseases (In Press).

Publications under review in peer-reviewed journals

Pulford CV*, Perez-Sepulveda BM, Canals R, Bevington JA, Bengtsson RJ[1], Wenner N, Rodwell EV, Kumwenda B, Zhu X, Bennett RJ, Stenhouse GE, De Silva PM, Webster HJ, Bengoechea JA, Dumigan A, Tran-Dien A, Prakash R, Banda HC, Alufandika L, Mautanga MP, Bowers-Barnard A, Beliavskaia AY, Predeus AV, Rowe W, Darby A, Hall N, Weill FX, Gordon MA, Feasey NA, Baker KS, Hinton JCD. (2020). The stepwise evolution of *Salmonella* Typhimurium ST313 responsible for bloodstream infection in Africa. *Currently under review at Nature Microbiology.*

Perez-Sepulveda BM*, Heavens D*, **Pulford CV***, Predeus AV, Low R, Webster H, Schudoma, C, Rowe W, Lipscombe J, Watkins C, Kumwenda B, Shearer N, Costigan K, Baker KS, Feasey NA, Hinton JCD, Hall N, The 10KSG consortium. (2020). An accessible, efficient and global approach for the large-scale sequencing of bacterial genomes. bioRxiv preprint. *Currently under review at Genome Biology.*

Kasumba IN, **Pulford CV**, Perez-Sepulveda BM, Sen S, Sayed N,PB Jasnehta, Livio S, Heavens D, Low R, Hall N, Roose A, Helen P, Farag T, Panchalingham S, Berkeley L, Nasrin D, Blackwelder WC, Wu Y, Tamboura B, Sanogo D, Onwuchekwa U, Sow SO, Ochieng JB, Omore R, Oundo JO, Breiman RF, Mintz ED, O'Reilly CE, Antonio M, Saha D, Hossain MJ, Mandomando I, Bassat Q, Alonso PL, Ramamurthy T, Sur D, Qureshi S, Zaidi AKM, Hossain A, Faruque ASG, Nataro JP, Kotloff KL, Levine MM, Hinton JCD, Tennant SM. (2020). Characteristics of *Salmonella* recovered from stools of children enrolled in the Global Enteric Multicenter Study. *Currently under review at Clinical Infectious Diseases.*

Bengtsson RJ, Dallman T, Jenkins C, Allen H, De Silva PM, Stenhouse GE, **Pulford CV,** Bennett RJ, Baker KS. (2020). Accessory genome dynamics and structural variation of *Shigella* from persistent infections. *Currently under review at mBio.*

International conferences and academic presentations

Invited presentation at the GBRU, Public Health England, London (2020)

Poster at the IMMEM, Dubrovnik, Croatia (2019)

Poster at the *Salmonella* Biology and Pathogenesis GRC, Boston, USA (2019)

Presentation at the GBRU, Public Health England, London, UK (2019)

Teaser presentation at the IIB Away Day, University of Liverpool, Liverpool, UK (2019)

Presentation at final year research showcase, University of Liverpool, Liverpool, UK (2019)

Presentation at Wellcome Trust Global Health Annual Scientific Meeting, Chester, UK (2018)

Invited presentation at the Microbiology Society ECR Conference, Birmingham, UK (2018)

Presentation at the GBRU, Public Health England, London, UK (2017)

Presentation at Wellcome Trust Global Health Research ASM, Chester, UK (2016)

Poster at American Society for Microbiology *Salmonella* ASM, Potsdam, Germany (2016)

Awards relating to this thesis

Best Talk Award for Biosciences at the University of Liverpool (2019)

Best PhD Final Year Presentation (2019)

Women in STEM Nova Award for Outstanding Contributions in Biology-£1,000 (2018)

Vice Chancellor Certificate of Excellence for Exemplary Academic Performance (2018)

John Lennon Memorial Award Towards Global and Public Health Research-£30,000 (2017)

Fee Bursary Award from Institute of Integrative Biology-PhD Tuition Fees (2016)

Contents

List of figures

List of tables

List of abbreviations

ACT	Artemis comparison tool
AIDS	Acquired immune deficiency syndrome
AMR	Antimicrobial resistance
BEAST	Bayesian evolutionary analysis by sampling trees
BEAUti	Bayesian evolutionary analysis utility
BH	Benjamini Hochberg
BLAST	Basic local alignment search tool
bp	Base pairs
C1	Clade 1
C2	Clade 2
C3	Clade 3
C4	Clade 4
C5	Clade 5
C6	Clade 6
C7	Clade 7
CI	Confidence interval
COG	Clusters of orthologous groups
Contig	Contiguous sequence
DI	Doherty Institute
DNA	Deoxyribonucleic acid
DRC	Democratic Republic of Congo
eBG	eBurst group
ESBLs	Extended spectrum beta-lactams
EUCAST	European committee on antimicrobial susceptibility testing
GTR	General time reversible
GWAS	Genome wide association study
HIV	Human immunodeficiency virus
HPD	Highest posterior distribution
Inc	Incompatibility group (plasmids)
iNTS	Invasive nontyphoidal *Salmonella*
IP	Institut Pasteur
iTOL	Interactive tree of life
kmer	DNA fragment of variable length k
L1	Lineage 1
L2	Lineage 2
L3	Lineage 3
LB	Lysogeny broth, Lennox
lrt	Likelihood ratio test
LSTM	Liverpool School of Tropical Medicine
MASCOT	Marginal approximation of the structured coalescent
Mbp	Million base pairs
MCMC	Markov chain monte carlo
MDR	Multidrug resistant
MLST	Multi locus sequence typing
MLVA	Multilocus variable number of tandem repeats analysis
MLW	Malawi-Liverpool Wellcome Clinical Research Programme
MNP	Multi-nucleotide polymorphisms

MRCA	Most recent common ancestor
NCBI	National Center for Biotechnology Information
NGS	Next generation sequencing
NTS	Nontyphoidal Salmonella
PCR	Polymerase chain reaction
PFGE	Pulse-field gel electrophoresis
PHE	Public Health England
QC	Quality control
qq plot	quantile-quantile plot
QUAST	Quality assessment tool for genome assemblies
RAxML	Randomised accelerated maximum likelihood
RDAR	Red, dry and rough (colony morphology)
RSPCA	Royal society for the prevention of cruelty to animals
SD	Standard deviation
SISTR	*Salmonella in silico* typing resource
SNP	Single nucleotide polymorphism
SPI	*Salmonella* pathogenicity island
SRA	Sequence read archive
ST	Sequence type
TAE	Tris-acetate-EDTA
TMAO	Trimethylamine *N*-oxide dihydrate
UBPE	Unité des bactéries pathogènes entériques
WGS	Whole genome sequencing
WHO	World Health Organisation
XDR	Extensively drug resistant
XLD	Xylose lysine deoxycholate

Chapter 1

Introduction

1.1 The emerging field of genomic epidemiology

1.1.1 Traditional bacterial typing methods

Bacterial typing methods have been used to inform epidemiological investigations since the 1920's, traditionally involving phenotypic classification of pathogens using labour intensive methods such as serotyping (Grimont and Weill, 2007) and phage typing (Ibrahim, 1969; Rabsch, 2007). Despite their initial success, these phenotypic tools have proven to lack the discriminatory power required in outbreak situations. Consequently molecular typing methods have been developed which can resolve complex population structures and monitor the genetic relatedness between bacterial pathogens (Deng et al., 2016). Early molecular typing tools included MultiLocus Variable number of tandem repeats Analysis (MLVA) (Best et al., 2007; Lindstedt et al., 2004) and Pulse-Field Gel Electrophoresis (PFGE) (Swaminathan et al., 2001). PFGE was pioneered in the mid 1990s by an organisation called PulseNet which played an instrumental role in improving the investigation of foodborne-disease outbreaks that were previously difficult to detect (Swaminathan et al., 2001). For example, eight times as many *Listeria* outbreaks were resolved in the five years following the introduction of PulseNet, compared with the twenty years prior (Cartwright et al., 2013). This routine use of molecular methods has also benefitted other foodborne bacteria such as *Salmonella*, and Shiga Toxin Producing *Escherichia coli* (STEC) (Tolar et al., 2019).

1.1.2 A brief history of whole genome sequencing

Meanwhile, a ground-breaking new technology was being developed called whole-genome sequencing (WGS), which would soon revolutionise the fields of public health, clinical care and scientific research (Heather and Chain, 2016). In 1977, Frederick Sanger developed a technique that involved the selective incorporation of chain-terminating dideoxynucleotides by deoxyribonucleic acid (DNA) polymerase during *in vitro* DNA replication, and was used to generate the world's first whole-genome sequence, that of a bacteriophage called φX174 (Sanger et al., 1977). Sanger's approach quickly became the most commonly used DNA sequencing technology worldwide (Heather and Chain, 2016). In subsequent years, the original Sanger method was improved by replacing phospho-radiolabelling of nucleotides with fluorometric based detection which facilitated the invention of increasingly automated WGS machines (Hunkapiller et al., 1991). The first generation of sequencing technologies were able to produce reads up to 1 kilobase (kb) in length, and multiple separate runs were required to analyse longer genomes. This shotgun sequencing approach was complemented

with computational processing, in which overlapping DNA fragments were assembled into longer contiguous sequences (contigs) (Anderson, 1981; Staden, 1979). Shotgun sequencing was used to complete the first bacterial genome sequences; *Haemophilus influenzae* (Fleischmann et al., 1995) and *Mycoplasma influenzae* (Fraser et al., 1995) in 1995, *Escherichia coli* (*E.* coli) in 1997 (Blattner et al., 1997) and *Salmonella enterica* in 2001 (McClelland et al., 2001). Despite being robust and simple, early sequencing technologies proved to be very expensive and labour intensive, and so their application on a larger scale was limited.

Subsequent advances in DNA sequencing were markedly different from the initial Sanger method, however have still utilised the same sequencing by synthesis approach (Heather and Chain, 2016). Specifically, researchers have utilised a novel luminescent method for measuring pyrophosphate synthesis from nucleotides which can be used in real time (Hyman, 1988; Nyrén and Lundin, 1985). Following ongoing development and refinement including the invention of the Flow Cell and bridge-amplification, this technology became much more high throughput, cost effective and accessible and came to be defined as next generation sequencing (NGS) (Shendure and Ji, 2008).

It is now possible to sequence complete microbial genomes at a price comparable to earlier subtyping methods, leading to an exponential increase in sequencing data over the past 20 years (Land et al., 2015). Broadly speaking, there are two types of whole genome sequencing technology; short read sequencers and long read sequencers (Quail et al., 2012). In recent years, the short-read sequencing industry has been dominated by Illumina (previously known as Solexa Ltd), whose high throughput platform can sequence reads up to 300 base pairs (bp) in length (Bennett, 2004). Specifically, the Illumina process uses clonal amplification and sequencing by synthesis. The process adds DNA bases into a nucleic acid chain and unique fluorescent signals emitted from each base are used to determine the order of the DNA sequence (Heather and Chain, 2016). Although Illumina sequencing is rapid, cheap and accurate; the method does not usually enable complete bacterial genome assembly, because reads are too short to reconstruct repetitive regions (Schatz et al., 2010). However, various bioinformatic approaches can be used to construct draft genome assemblies, which are widely used in research and public health (Ashton et al., 2016). Long read sequencing can be used instead-of or as-well-as short read sequencers to resolve complex bacterial genomes,

largely pioneered by companies such as Pacific Biosciences (Rhoads and Au, 2015) and Oxford Nanopore Technologies (Clarke et al., 2009; Eisenstein, 2012). In particular, Oxford Nanopore's MinION machine, has generated a great deal of excitement in recent years, as the first mobile DNA sequencer which is able to be deployed rapidly and in-field, unlocking the power to decentralise WGS (Jain et al., 2016).

1.1.3 Bioinformatic approaches used to process sequencing data

Alongside the huge availability of WGS data, countless bioinformatic tools have been developed to aid genomic analysis (Deng et al., 2016). A fundamental requirement for analysing WGS data, is the ability to assemble short reads into more complete genomes. Most short read *de novo* assembly pipelines are based on de Bruijn graph strategies, involving breaking reads into shorter DNA fragments of variable length k (kmers), after which a graph of perfectly overlapping kmers is constructed (Compeau et al., 2011; Miller et al., 2010). A major challenge for short read assembly pipelines is reconstruction of repetitive DNA regions called tandem repeats, which are sequences composed of several DNA fragments lying one after another (Fan and Chu, 2007). Common genome assemblers such as SPAdes (Bankevich et al., 2012), Velvet (Zerbino, 2010), and ABySS (Jackman et al., 2017), used paired end information to overcome issues with tandem repeats. Short read data will typically result in draft genome assemblies, composed of multiple contigs (Parkhill, 2013). Hybrid assembly pipelines which combine short and long read data such as Unicycler can be used to produce a complete genome (Wick et al., 2017).

Alternatively, reference-based assembly can be used which involves mapping short read data against a complete, genetically similar and preferably high-quality reference genome (Deng et al., 2016). The mapping process involves building an index for the reference genome, which is used to identify the genomic positions of each read. There are various techniques which may be used to build the index such as hash-tables and an algorithm known as Burrows Wheeler Transform (Hatem et al., 2013). In particular, Burrows Wheeler Transform is a very efficient indexing technique which was extended to a newer data structure called FM-index to support instances where a single read matches multiple locations across the genome (Ferragina and Manzini, 2000). Tools which use the Burrows Wheeler Transform include Bowtie (Langmead and Salzberg, 2012), BWA (Li and Durbin, 2009) and SOAP2 (R. Li et al., 2009). Although reference-based mapping techniques do not take into consideration regions

which may be missing from the reference genomes, the approach is relatively rapid, computationally inexpensive and widely used in the scientific community.

Short read mapping is an ideal tool for identifying genomic regions which differ between a query sequence and a reference and may be used to detect single nucleotide polymorphisms (SNPs) used for phylogenetic inference through variant calling. Typical tools used for the mapping and variant-calling process have been implemented through programs such as SAMtools (H. Li et al., 2009; Li, 2011), GATK (McKenna et al., 2010), Freebayes (Garrison and Marth, 2012), VarScan (Koboldt et al., 2009) and Cortex (Iqbal et al., 2012). Cortex is unique in that it is a de Bruijn graph-based tool that detects SNPs by directly loading reads from multiple samples into the same graph (Iqbal et al., 2012). As WGS has become more accessible, some dedicated SNP-calling pipelines have begun to emerge for bacterial isolates such as Snippy (https://github.com/tseemann/snippy), CFSAN (Davis et al., 2015), NASP (Sahl et al., 2016) and PHEnix (https://github.com/phe-bioinformatics/PHEnix). Generally, these SNP-calling pipelines involve an external mapping tool and variant caller, followed by a series of filtering steps to remove low-quality SNPs. A limitation of the SNP-calling process is that there may be an accumulation of false positive SNPs amongst closely related bacterial isolates, with new pipelines such as BactSNP aiming to overcome this issue (Yoshimura et al., 2019).

1.1.4 An overview of phylogenetic methods

Phylogenetics are used to infer the evolutionary history and structure of bacterial populations (e.g. Nightingale, Windham and Wiedmann, 2005; Holt *et al.*, 2012; Mather *et al.*, 2013; Wong *et al.*, 2016). A phylogenetic reconstruction usually involves applying an evolutionary model to an alignment of core genome SNPs which vary against a reference genome. Highly recombinant regions are excluded from the alignment as they may impair accurate phylogenetic reconstruction (Croucher et al., 2014). The main reason for these inaccuracies is because recombination results in a large number of SNPs compared to the reference genome; because phylogenetic methods have an underlying assumption of vertical evolution, recombination tends to inflate branch lengths, although generally topology is not affected (Hedge and Wilson, 2014). The extent of recombination varies greatly among bacterial species, for example *Escherichia coli* has a very low recombination

rate and *Neisseria gonorrhoeae* has a very high recombination rate (Pérez-Losada et al., 2006).

There are multiple approaches used for phylogenetic inference, including distance-based, maximum parsimony, maximum likelihood and Bayesian (Svennblad et al., 2006). Maximum-likelihood methods use a hill climbing algorithm to identify the tree that is most likely to have produced the observed data, given a model of evolution (Felsenstein, 1981). Specifically, the evolutionary model is used to estimate base frequencies and substitution rates and ranges from simple (e.g. Jukes Cantor 69) to more complex (e.g. General Time Reversible (GTR) plus gamma distribution). Maximum likelihood approaches are commonly executed through programs such as RAxML (Randomised Accelerated Maximum Likelihood) (Stamatakis, 2014) or IQ-tree (Nguyen et al., 2015). Note that maximum parsimony approaches also use a hill climbing algorithm to search through tree topologies, but using a much simpler model of evolution (Giribet, 2007). In contrast, Bayesian algorithms use Markov Chain Monte Carlo (MCMC) to average over tree space and achieve the highest posterior probability given the observed data and a set of prior assumptions. The most widely-used Bayesian phylogenetic software is BEAST (Bayesian Evolutionary Analysis by Sampling Trees) (Drummond and Rambaut, 2007), although other programs such as MrBayes are also available (Ronquist et al., 2012). Although Bayesian methods are generally considered to be less user friendly and more computationally intense than maximum likelihood methods, the approach has incredible power and facilitates the study of the spatio-temporal process (Lau et al., 2015). Thus, Bayesian approaches allow key epidemiological quantities such as contact tracing latent periods, population expansion and migration rates to be estimated, which are critical to risk assessments and disease control.

1.1.5 Deriving phenotypic information from bacterial genomes

Bacterial genomes can also be used to monitor trends in antimicrobial resistance (AMR), a global health problem which results in hundreds of thousands of deaths every year (O'Neill, 2019). Traditionally AMR has been detected by measuring the inhibitory effects of a specific antimicrobial agent on bacterial growth using phenotypic methods (Brown et al., 2015). However, AMR can also be predicted by identifying resistance conferring genes and mutations in bacterial genomes (Hendriksen et al., 2019). By 2019, there were at least 47 freely accessible bioinformatic resources for detection of AMR determinants in genome

sequence data including databases such ResFinder (including PointFinder), AMRFinder and CARD (Alcock et al., 2019; Feldgarden et al., 2019; Zankari et al., 2017, 2012). AMR detection tools may be contig-based such ABRicate and staramr or may use short reads such as SRST2 and ARIBA, (Hunt et al., 2017; Inouye et al., 2014). Previously, whole genome-based analysis accurately predicted the AMR phenotypes of 89.8 % (1007/1122) of *Salmonella enterica* serovar Typhimurium isolates, with 83% sensitivity and 96% specificity (Mensah et al., 2019). By studying the diversity, temporal and phenotypic relatedness of resistance profiles, studies have been able to infer probable sources of AMR in bacterial pathogens, which can then be used to inform policy choices (Mather et al., 2012).

A powerful tool that can be used to investigate associations between genotype and phenotype is a genome wide association study (GWAS) (Lees and Bentley, 2016). GWAS approaches have successfully identified resistance-conferring variants in *Mycobacterium tuberculosis* (Desjardins et al., 2016). In addition to studying AMR, GWAS' have also been used to study host resistance in *Salmonella* Pullorum (Li et al., 2019) and the genetic basis of increased virulence in *Listeria monocytogenes* (Maury et al., 2016) and *Staphylococcus aureus* (Young et al., 2019). However, the clonality of bacterial populations represents a major hurdle for GWAS methods. Specifically, when comparing lineages with different phenotypes, all variants that separate the lineages appear to be associated with a phenotype even though there is no causal link (Lees and Bentley, 2016). Thus, various models have been developed to incorporate a correction for population structure (Lees et al., 2019, 2018).

1.1.6 Defining genomic epidemiology

WGS allied with an expanding repertoire of bioinformatic tools has been a cornerstone in modern science. The term genomic epidemiology has increasingly been used to describe the process of using DNA sequence data to determine temporal, geographical and evolutionary trends in microbial pathogens (Kan et al., 2018; Mather et al., 2015). Genomic epidemiology has proven to be instrumental in pathogen surveillance and has been particularly used for monitoring enteric bacteria (Hawken and Snitkin, 2019).

In recent years, routine WGS of *Salmonella* isolates from foodborne disease outbreaks has being adopted by public health authorities in high income countries, including the United

Kingdom and the United States of America (Ashton et al., 2016; Ribot et al., 2019). These data have been complemented by global efforts to increase sequencing in low-middle income countries, for example that driven by the 10,000 *Salmonella* genomes project (Perez-Sepulveda et al., 2020). Sharing of WGS data has been facilitated by publicly available repositories such as those hosted by the NCBI (Leinonen et al., 2011). Organism specific databases also exist such as EnteroBase, which consists of a combination of data sourced from the National Center for Biotechnology Information (NCBI) sequence read archives (SRA) and user uploads (Zhou et al., 2020). Indeed, there are 258,817 *Salmonella* genomes alone on EnteroBase at the time of writing (https://enterobase.warwick.ac.uk/, accessed 1 August 2020), a figure which has doubled since 2018 (**Figure 1.1**)(Alikhan et al., 2018). At the moment, there are more than twice as many *Salmonella* genomes in the SRA as any other bacterial pathogen (Moustafa et al., 2020).

As the cost of WGS continues to decrease, and more user-friendly bioinformatic approaches begin to emerge, genomics has become a viable and advanced solution for epidemiological investigation and intensive scientific research for many bacterial species, including the globally relevant pathogen *Salmonella*.

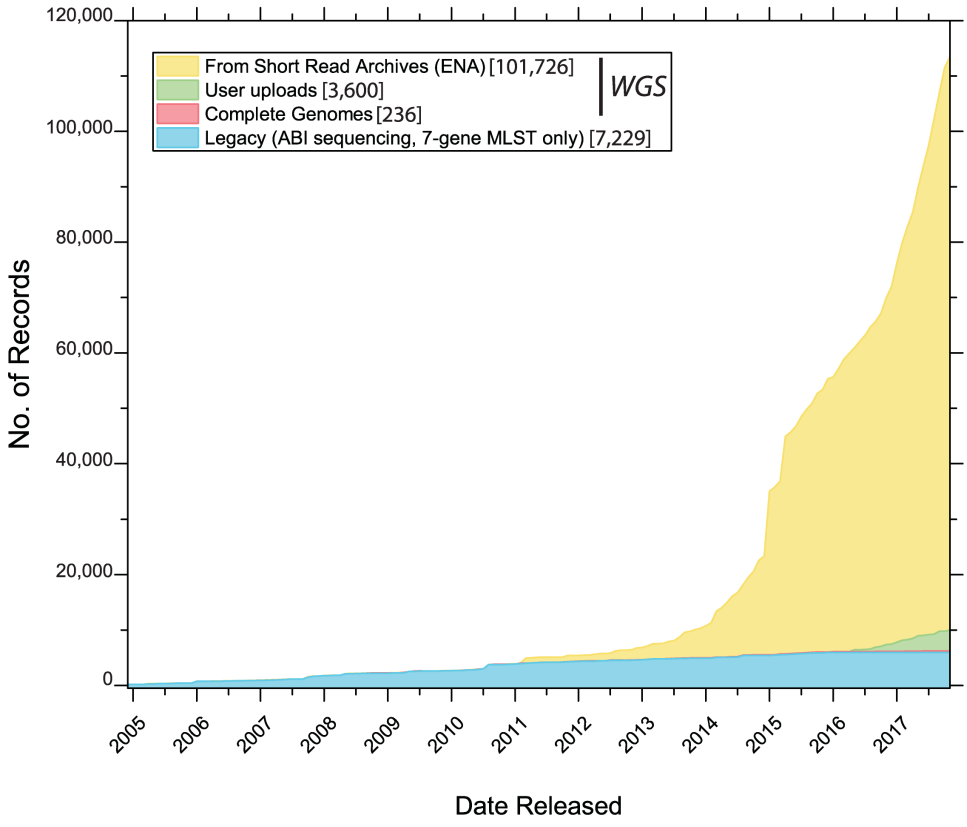


Figure 1.1 The increasing number of *Salmonella* genomes available on EnteroBase 2005-2018

EnteroBase is a user-friendly online resource for analysing and visualising genomic variation within enteric bacteria. The platform has been automatically assembling *Salmonella* genomes from sequencing reads that have been submitted to the SRA (yellow) or directly uploaded to EnteroBase by users (green). EnteroBase also contains a small number of complete genomes (pink) and a number of entries with legacy MLST genotypes (blue). **(Figure from Alikhan *et al.,* 2018).**

1.2 Subtyping schemes for *Salmonella*

1.2.1 The complexity of *Salmonella* nomenclature

The genus *Salmonella* consists of rod-shaped bacteria belonging to the Enterobacteriaceae family, and more broadly the class Gammaproteobacteria (Tindall et al., 2005). *Salmonella* taxonomy is complex, and current usage often combines many classification systems that inconsistently discriminate between species, subspecies, serovars and sequence types (Brenner et al., 2000). **Figure 1.2** summarises the complexity of the *Salmonella* classification system.

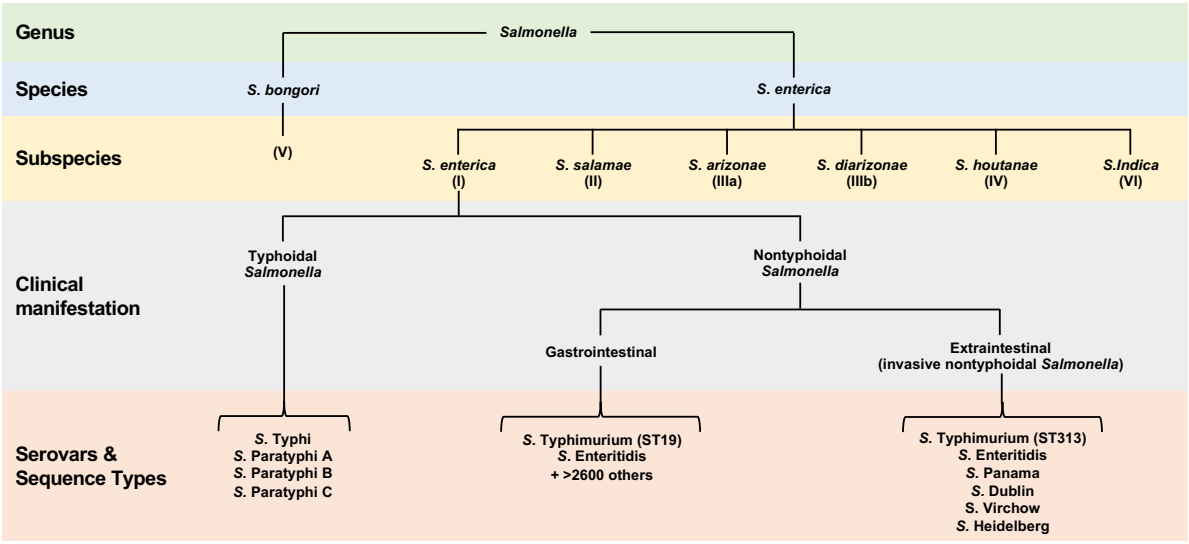


Figure 1.2 A summary of *Salmonella* nomenclature

The complexity of *Salmonella* nomenclature can be broken down into genus (green), species (blue), subspecies (yellow) clinical manifestation (grey), serovars and sequence types (orange). Note that a comprehensive list of all *Salmonella* serovars and sequence types is not provided for extraintestinal serovars, rather common serovars are shown. **(Figure adapted from Langridge *et al.,* 2013).**

1.2.2 History of *Salmonella* subtyping

Historically, *Salmonella* species were defined on the basis of serological identification of O (somatic) and H (flagella) antigens (Kauffmann, 1966). If this one species-one serotype concept was still in use today, there would be over 2,600 different species of *Salmonella* (Issenhuth-Jeanjean et al., 2014). However, in 1973 a method known as DNA-DNA hybridisation was developed which demonstrated that multiple serovars could be linked at the species level, in a defining moment for *Salmonella* taxonomy (Crosa et al., 1973). Subsequently, two *Salmonella* species were identified, and initially named *S. bongori* (which still stands today) and *S. choleraesuis.* The name *S. choleraesuis* was originally chosen because it appeared on the Approved List of Bacterial Names, but caused substantial

confusion owing to the fact it was used to classify both the species and the serovar (Brenner et al., 2000). Thus, in 1986 the International Sub-Committee in Systemic Bacteriology for Enterobacteriaceae unanimously recommended the species designation for *S. choleraesuis* be changed to *S. enterica* (Penner, 1988) a name coined by Kauffman and Edwards in 1952 (Kauffmann and Edwards, 1952).

1.2.3 Species-level classification

There are currently two recognised *Salmonella* species; *S. enterica* and *S. bongori* (Brenner et al., 2000), which are thought to have diverged from a recent common ancestor with *Escherichia coli* (*E. coli*) over 100 million years ago (Battistuzzi et al., 2004; Desai et al., 2013). The genomes of *Salmonella* and *E. coli* still display extensive regions of synteny; whilst differing in the presence of large genomic regions known as the *Salmonella* Pathogenicity Islands (SPIs) which are crucial for the intracellular lifestyle of *Salmonella.* The species *S. bongori* is primarily found in cold-blooded animals and lacks SPI-2 which is required for optimal replication within mammalian macrophages. *S. bongori* is therefore considered to be an evolutionary intermediate between *E. coli* and *S. enterica* (Fookes et al., 2011).

S. *enterica* is further divided into six subspecies, which reflect previously defined subgenera and differ both biochemically and by genomic content (Ryan et al., 2017). Specifically, the *Salmonella* subspecies are; *S. enterica* subsp. *enterica* (I), *S. enterica* subsp. *salamae* (II), *S. enterica* subsp. *arizonae* (IIIa), *S. enterica* subsp. *diarizonae* (IIIb), *S. enterica* subsp. *houtenae* (IV) and *S. enterica* subsp. *indica* (VI). Henceforth, the term *Salmonella* will be used to refer to the *S*. *enterica* species and the *S*. *enterica* designation will be used to refer to the *S*. *enterica* subspecies *enterica* alone, unless stated otherwise.

1.2.4 *Salmonella* serovar designations

The *Salmonella* subspecies can be further classified into serovars which are ecologically, phenotypically and genetically diverse (Winfield and Groisman, 2004). The serotyping scheme relies on agglutination of a bacterial strain with specific antisera that has been raised against well-characterised *Salmonella* strains. An antigenic formula is then assigned to the strain based on the identification of O and H antigenic variants (Grimont and Weill, 2007). There are at least 46 O-antigens and 14-H antigens that have been identified, which in combination characterise over 2,600 different serovars (Issenhuth-Jeanjean et al., 2014).

Serotyping was designed as an epidemiological typing tool to aid differentiation among *Salmonella* variants in diagnostic and public health settings. However, significant limitations are associated with this traditional method including low throughput methodology, high expense, the requirement for expert personnel and the over 150 specific antisera required to be generated through immunising rabbits (McQuiston et al., 2004). In the current genomic-era of pathogen surveillance, several *in silico* platforms utilising WGS data to predict serotype have now been developed (Ashton et al., 2015; Yoshida et al., 2016; Zhang et al., 2015). This computational serotyping approach involves detecting SNPs within flagella antigens (Yoshida et al., 2016). An underlying assumption of serotyping is that serovar designations reflect genetic relatedness, which holds true in most instances within the *Salmonella* genus. Notable exceptions are in the serovars; Java, Paratyphi C and Choleraesuis. (Achtman et al., 2012).

Approximately 60% of all of the recognised *Salmonella* serovars belong to *S. enterica* (I) (Grimont and Weill, 2007). Phylogenetic analysis of *S. enterica* serovars has shown evidence of at least two sub-populations (clade A and clade B) with distinct genome characteristics including host-specific virulence gene repertoires, suggesting differences in transmission characteristics and ecological niches occupied by the groups (den Bakker et al., 2011; Falush et al., 2006; Parsons et al., 2015). Although *S. enterica* clade B serovars have been associated with an increased ability to cause clinically systemic infections, the majority of serovars which are able to cause human disease belong to *S. enterica* clade A (Worley et al., 2018). Interestingly, the distribution of disease caused by the ~1,500 *S. enterica* serovars is highly skewed. Over the past decade for example, just six nontyphoidal *Salmonella* serovars have been responsible for ~70% of cases of human disease in Europe and the USA, namely Enteritidis, Typhimurium, monophasic Typhimurium 1,4,[5],12:i:-, Infantis, Newport and Javiana (Centers for Disease Control and Prevention, 2018; European Food Safety Authority and European Centre for Disease Prevention and Control (EFSA and ECDC), 2019).

1.2.5 Phage typing

To complement and enhance serological classifications, phage typing was introduced as a rapid, accurate and cheap tool for epidemiological use (Ibrahim, 1969). This classification system involves differentiating *Salmonella* into a number of phage types based on their pattern of susceptibility to killing by a specific set of bacteriophages (viruses that can infect

bacteria) (Rabsch, 2007). Like antibodies in sera, phages recognise receptors on the bacterial cell surface, and consequently phage typing correlates well with serology (Ibrahim, 1969). Since the 1990s, phage typing proved to be an important tool for grouping strains at the sub-serovar level and has successfully been used for source attribution and outbreak investigations within the public health setting (Baggesen et al., 2010).

1.2.6 Multi Locus Sequence Typing

Though traditional typing methods have played an important role in *Salmonella* disease epidemiology, molecular methods have provided a new level of discriminatory resolution and allowed evolutionary relatedness to be incorporated into the classification process (Alikhan et al., 2018). Multi Locus Sequence Typing (MLST) is a system that has been introduced to support (and possibly replace) previous *Salmonella* nomenclature (Achtman et al., 2012). The scheme was initially developed to discriminate *S.* Typhi clusters (Kidgell et al., 2002), and has subsequently been used throughout the *S. enterica* subspecies. Specifically, MLST involves assessment of the conservation of seven housekeeping gene fragments, specifically *thrA, purE, sucA, hisD, aroC, dnaN* and *hemD* (Maiden et al., 2013). The seven genes were selected based on the fact they are scattered around the chromosome, are flanked by genes of known function and that neither the gene nor the flanking genes are likely to be under diversifying selection (Kidgell et al., 2002). For a given isolate, the sequence for each of these genes is determined and each known allelic variant is assigned a number based on the chronology of the discovery of that variant. The combination of the seven alleles is then assigned a Sequence Type (ST) (Maiden et al., 2013). STs which share 6/7 alleles can be clustered into 139 genetic populations known as eBurst Groups (eBGs), which correlate to a certain extent with serovar designations (**Figure 1.3**). For example, the serovar *S.* Typhimurium corresponds to eBG 1 and contains multiple sequence types including ST19 which is responsible for foodborne gastrointestinal infection globally (Achtman et al., 2012)

Traditionally, the allelic sequences of the seven housekeeping genes were obtained using Polymerase Chain Reaction (PCR) amplification followed by Sanger Sequencing, but more recently WGS data has been used to derive MLST status (Alikhan et al., 2018). As a result, MLST schemes with an even greater power of resolution have been introduced, largely pioneered through the online platform EnteroBase. *Salmonella* schemes include; ribosomal

MLST (51 loci), core genome MLST (3,002 loci) and whole genome MLST (21,065 loci) (Alikhan et al., 2018). In particular core genome MLST has been used for phylogenetic inference, providing an alternative and reference-free methodology to SNP-based approaches.

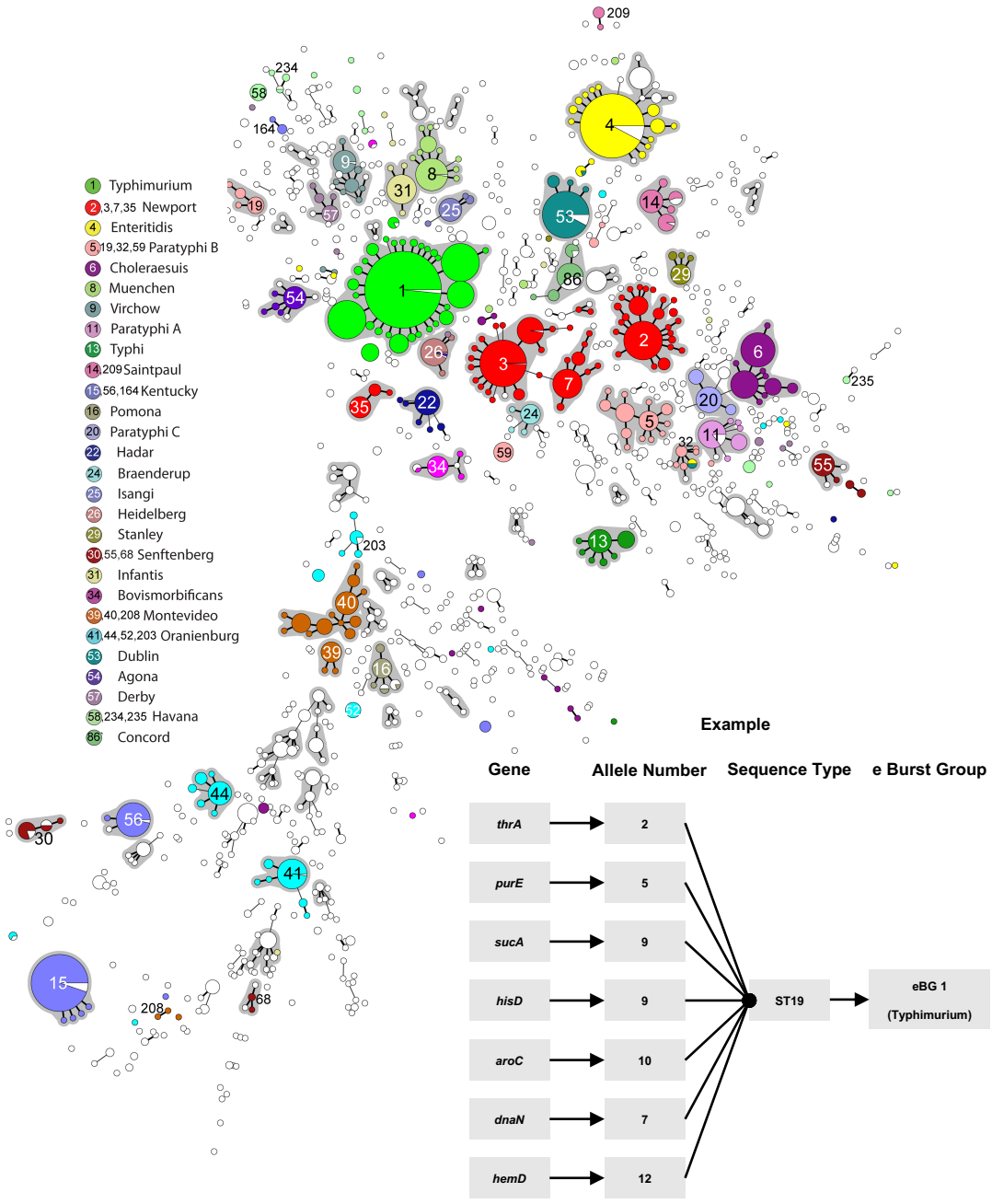


Figure 1.3 Genetic population structure of *Salmonella* based on MLST data

Minimal spanning tree of 4,257 isolates of *S. enterica* available on EnteroBase. Each circle corresponds to a sequence type. Grey shading is used to indicate eBGs and cluster labels refer to eBG. Serovars associated with the most isolates in each eBG is indicated with a colour and a number. An example of the MLST scheme is also provided showing the MLST genes, allele number, sequence type, serovar and corresponding eBG **(Figure adapted from Achtman *et al.*, 2012).**

1.3 Host-range

1.3.1 Differences between host-restricted, adapted and generalist *Salmonella*

Salmonella has a very broad host-range at the genus level, being able to survive in a wide range of niches, however individual subspecies and serovars lie on a spectrum of host-adaptation (Bäumler and Fang, 2013). A small number of *Salmonella* serovars are considered to be host-restricted, having evolved to colonise only a very limited range of hosts. For example, *S.* Typhi, *S.* Sendai and *S.* Paratyphi A, B and C are highly adapted to humans (Uzzau et al., 2000). Reservoirs for these specialised serovars include water, vegetables, seafood and raw eggs (Gal-Mor et al., 2014). Host-restricted serovars are also apparent in other animal species including *S.* Gallinarum in chickens, *S.* Typhisuis in swine, *S.* Abortusovis in sheep and *S.* Abortusequi in horses (Shivaprasad, 2000; Uzzau et al., 2000). Other serovars of *Salmonella* are regarded as host-adapted, and are usually associated with a particular animal host (Kingsley and Bäumler, 2000). Examples include *S.* Dublin in cattle and *S.* Choleraesuis in pigs, both of which can also cause disease in humans (Uzzau et al., 2000).

However, the majority of *Salmonella* serovars can infect a wide variety of hosts including humans, wild and domestic animals and the environment, they are therefore described as host-generalists (Langmead and Salzberg, 2012; Li and Durbin, 2009). These serovars are mainly transmitted to humans through the food-chain including beef, pork, poultry and contaminated raw eggs (Ferrari et al., 2019). Although host-generalist *Salmonella* can infect a wide range of hosts, serovars are often linked with certain reservoirs. For example, poultry and contaminated chicken eggs have played a central role in the dissemination of *S.* Enteritidis into humans, causing multiple outbreaks over the past several decades (Betancor et al., 2010; Braden, 2006; Dallman et al., 2016; Pijnacker et al., 2019). A recent metadata analysis of *Salmonella* serovars in animal-based foods additionally highlighted links between *S.* Anantum in beef, *S.* Weltevreden in seafood, and *S.* Typhimurium in swine (Ferrari et al., 2019). Indeed, pigs have been regarded as the most likely reservoir for a pandemic of monophasic ST34 *S*. Typhimurium variant, *S*. I:4,[5],12:i:− (Antunes et al., 2011; Hopkins et al., 2010). In addition to livestock, wild animals may also play a role in dissemination of *Salmonella* serovars such as wild passerines (Mather et al., 2016) and reptiles (Whiley et al., 2017).

1.3.2 *Salmonella* in reptiles

Both wild and domestic reptiles are natural asymptomatic reservoirs for host-generalist *Salmonella* (Hoelzer et al., 2011; Warwick et al., 2001). The reported prevalence of *Salmonella* carriage in reptiles is highly variable amongst the literature, ranging from 14% to 100% with inconsistent results reflecting intermittent faecal shedding patterns (Goupil et al., 2012; Lukac et al., 2015). Historically, non-enterica *Salmonella* subspecies have been associated with carriage in cold-blooded animals such as reptiles and amphibians, but rarely in humans (Bäumler and Fang, 2013; Clancy et al., 2016; Nair et al., 2014; Schroter et al., 2004). However, reptiles are also capable of carrying serovars belonging to *S. enterica,* a species that is responsible for causing the majority (99%) of human cases of salmonellosis (Anon, 2013; Eng et al., 2015; Kanagarajah et al., 2018; Murphy and Oshin, 2015; Whiley et al., 2017).

Salmonella can be transmitted from reptiles to humans in several ways resulting in reptile associated salmonellosis (RAS). RAS occurs worldwide and is recognised as an emerging disease of global public health significance (Whiley et al., 2017). In the USA, it is estimated that around 6% of human *Salmonella* cases are attributed to reptiles each year (Mermin et al., 2004) and in South West England, 27.4% of *Salmonella* cases in children under age five were linked to reptile exposure (Murphy and Oshin, 2015). The latter study also demonstrated that RAS was more likely to lead to systemic disease resulting in hospitalisation than non-RAS (Murphy and Oshin, 2015).

1.4 The relationship between host-specificity and clinical syndrome

Generally, there is an inverse relationship between the breadth of host range and clinical disease severity, with host-restricted *Salmonella* tending to cause systemic disease and host-generalist *Salmonella* tending to cause gastrointestinal disease (**Figure 1.4**) (Kingsley and Bäumler, 2000). The association between the various clinical outcomes caused by different serovars was described by Jones *et al.* who studied a large dataset of *Salmonella* collected over ten years from the USA. The study made the conclusion that disease outcome is a function of serovar and that understanding this may be the key to a more general understanding of invasiveness (Jones et al., 2008). Within humans, *Salmonella* causes a huge range of clinical disease syndromes, including enteric fever, gastroenteritis, sepsis, meningitis and focal infections (Stanaway et al., 2019a, 2019b). Because of the huge number of serovars and the various clinical disease outcomes, clinicians have found it useful to broadly classify *Salmonella* as typhoidal or nontyphoidal.

1.4.1 Typhoidal *Salmonella*

Typhoidal *Salmonella* is restricted to human hosts and presents as a systemic infection. The most recent estimate for disease burden is 14.3 million cases and 135,900 fatalities according to the "global burden of diseases, injury and risk factors" study in 2017 (Stanaway et al., 2019b). However, earlier estimates in 2010 had predicted 26.9 million cases and approximately 200,000 typhoid related deaths (Buckle et al., 2012). Only a handful of serovars are known to cause the febrile bacteraemia syndrome associated with Typhoid fever including *S.* Typhi and *S.* Paratyphi A, B and C. These Typhoid and Paratyphoid infections most commonly affect countries which have a poor water quality and modest sanitation, especially south Asia, south East Asia and sub-Saharan Africa (Crump et al., 2004; Mogasale et al., 2014).

1.4.2 Nontyphoidal *Salmonella*

The remaining serovars are considered to be nontyphoidal *Salmonella* (NTS). NTS infections most commonly result in a self-limiting gastrointestinal disease with low fatality and occur worldwide, with links to industrialised food production (Ricke et al., 2015). In 2010, NTS disease was estimated to be responsible for 93.8 million gastroenteric illnesses and 155,000 deaths worldwide (Majowicz et al., 2010). However, in 2017 the "global burden of diseases, injuries and risk factors study" updated these original approximations, estimating that the

worldwide disease burden of NTS was 95.1 million cases of enterocolitis and 50,771 deaths worldwide (Stanaway et al., 2019a).

1.4.3 Invasive nontyphoidal *Salmonella*

A recent systematic review established that nontyphoidal *Salmonella* was also the most frequently-isolated pathogen found in a large cohort of patients suffering from community-onset bloodstream infections in both Africa and Asia between 2008 and 2018 (Marchello et al., 2019). The associated disease, invasive nontyphoidal *Salmonella* (iNTS), poses a significant burden to public health globally, resulting in 535,000 cases and 77,500 deaths in 2017 (Stanaway et al., 2019a). Incidence is particularly high in sub-Saharan Africa, resulting in 34.5 cases per 100,000 person years (Stanaway et al., 2019a). Although iNTS is comparatively uncommon in other parts of the world, the disease has also been described in Southeast Asia, albeit in a relatively smaller patient cohort (103 iNTS patients in an infectious disease hospital in Southern Vietnam between 2008 and 2013) (Phu Huong Lan et al., 2016). Specific host-factors are considered to increase susceptibility of humans to iNTS infection; including Malaria and malnutrition in infants and Human Immunodeficiency Virus (HIV) in adults (Feasey et al., 2012; Gordon, 2008). Indeed, recurrent iNTS disease was added to the United States Centres for Disease Control and Prevention case definition for HIV in 1987 and by 1990 NTS had been confirmed as a common Acquired Immune Deficiency Syndrome (AIDS)-related pathogen in sub-Saharan Africa (Feasey et al., 2012; Gilks et al., 1990).

1.4.4 Human-human transmission of NTS and iNTS

Although the majority of NTS infections are foodborne and are ultimately animal-derived, there is mounting evidence that several key pathovariants spread via human-human transmission. For example, a recent study focused on novel biphasic *S.* Typhimurium ST34 variants causing iNTS disease in Vietnam, found little evidence of zoonotic transfer between co-locating animals and humans (Mather et al., 2018). Similarly, an analysis in Kenya showed that there was no significant relationship between NTS isolated from humans and those isolated from animals, food or environmental samples within or near iNTS patient homes (Kariuki et al., 2002). Further investigations in Kenya have revealed that NTS isolated from the bloodstream of cases was more comparable to that found in asymptomatic human carriers than that found in environmental sources including food, sewers, soil and animals

(Kariuki et al., 2006). Comparably, in Malawi different serovars of NTS were isolated from patients with iNTS disease and livestock within the same households (Post et al., 2019). In an industrialised setting, WGS was used to show that the major source of human NTS infection and the AMR associated with those infections was unlikely to be local animal populations (Mather et al., 2013).

1.4.5 Gene degradation

Pathovariants of iNTS display genetic signatures which suggest adaptation to an invasive and host-adapted lifestyle. It was initially believed that *Salmonella* evolution was driven primarily by the acquisition of virulence genes that conferred some sort of selective advantage to the bacteria (Bäumler et al., 1998). However, WGS-based data has revealed a much more complex evolutionary perspective, specifically that host-adapted, extraintestinal *Salmonella* variants have evolved independently multiple times from gastrointestinal ancestors and show a greater degree of genomic degradation in comparison with host-generalist *Salmonella* (Bäumler and Fang, 2013; Key et al., 2020; Zhou et al., 2018). Genomic degradation is a term used to describe an accumulation of loss-of-function mutations and deletions across a genome. For instance, functions required for promoting colonisation of the inflamed mammalian gut are often lost in invasive, host-adapted serovars in a series of gene pseudogenisation events (Nuccio and Bäumler, 2014; Okoro et al., 2015).

Examples of gene pseudogenisation are apparent across the *S. enterica* species, with host-generalist serovars containing far fewer pseudogenes than host-restricted serovars. A deeper insight into the evolutionary progression towards host-adaptation can be seen by looking at genomic comparisons between ancient and contemporary *S. enterica* isolates (Key et al., 2020). A recent study has shown how ancient *S. enterica* were host-generalist and experienced convergent evolution of pseudogenisation that accompanied downstream host adaptation (Key et al., 2020). Indeed, the number of pseudogenes in the host-generalist *S.* Typhimurium strain LT2 (*n* = 25) is significantly less than in systemically adapted host-restricted serovars such as *S.* Gallinarum (*n* = 309), *S.* Typhi (*n* = 204), *S.* Dublin (*n* = 177), *S.* Paratyphi A (*n* = 173 pseudogenes), *S.* Paratyphi C (*n* = 152 pseudogenes) or *S.* Choleraesuis (*n* = 151 pseudogenes) (Betancor et al., 2012; Chiu et al., 2005; Liu et al., 2009; McClelland et al., 2004; Parkhill et al., 2001; Thomson et al., 2008). However, there are notable exceptions where important iNTS pathovariants display little evidence of genome

degradation including a multidrug resistant (MDR) sub-group of *S.* Typhimurium ST34 in Vietnam (Mather et al., 2018).

The invasiveness of some *Salmonella* serovars can be assessed experimentally with animal infection models (Tsolis et al., 2011). However, given that systemic pathovariants are often linked with host-adaptation, there have been concerns regarding the use of animal models. Several previous analysis have demonstrated that cellular and animal models fail to distinguish between levels of invasiveness in extraintestinal and gastrointestinal *Salmonella* variants (Lacharme-Lora et al., 2019; Van Puyvelde et al., 2019). In theory, the extent of genome degradation in *S. enterica* pathovariants can indicate the extraintestinal potential of the pathogen. Initial attempts to measure pseudogenisation between bacterial genomes have been labour intensive, relying on the manual annotation of truncations, deletions and frameshifts in genes (Langridge et al., 2015; Nuccio and Bäumler, 2014).

A more recent approach involved the calculation of an invasiveness index, a value that represents the extent of genome degradation and diversifying selection specific to invasive serovars using a set of top extraintestinal predictor genes (Wheeler et al., 2018). The approach specifically uses a random forest classifier and delta bitscore functional variant calling to place *Salmonella* strains on spectrum from invasive to gastrointestinal. The invasiveness index has been validated using multiple *Salmonella* serovars and the approach has also been used to discriminate between *S.* Enteritidis and *S.* Typhimurium lineages associated with extraintestinal infections in sub-Saharan Africa (Van Puyvelde et al., 2019; Wheeler et al., 2018).

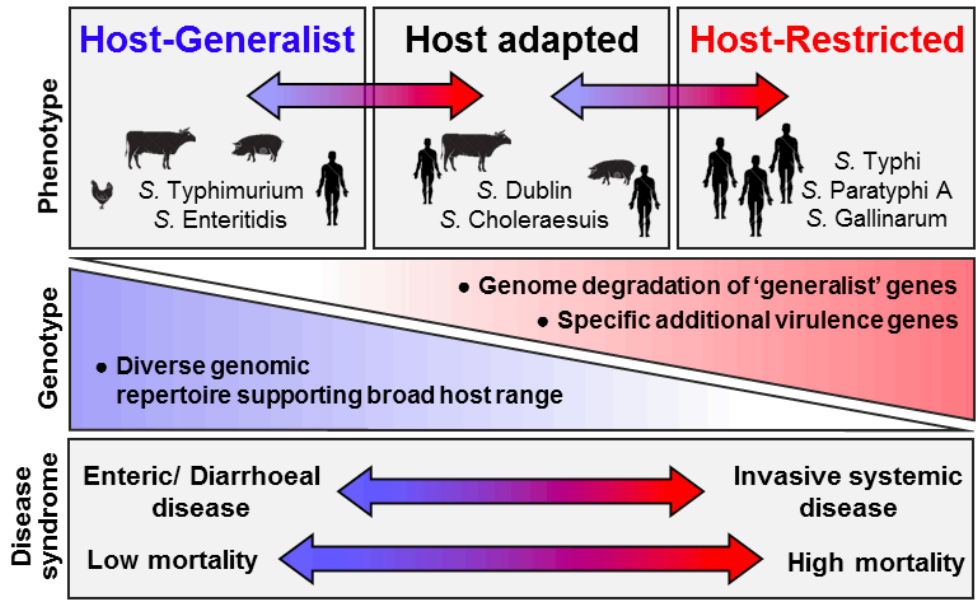


Figure 1.4 The host adaptation and disease syndrome paradigm

Infographic overview of the relationship between host-specificity and clinical disease syndrome for *S. enterica.* (**Figure provided by Siân Owen**)

1.5 An introduction to *S.* Typhimurium

1.5.1 *S.* Typhimurium in sub-Saharan Africa

The most common iNTS serovars in sub-Saharan Africa have been *S.* Typhimurium and *S.* Enteritidis (Uche et al., 2017). This is exemplified by a recent cross-sectional study focused on iNTS samples collected from children in the Democratic Republic of Congo, which showed that 68.6% of isolates were *S.* Typhimurium, 30.4% were *S.* Enteritidis and only 1% were classified as other serovars (Phoba et al., 2020). A high proportion of *S.* Typhimurium infections has been seen in several countries across sub-Saharan Africa, with meta-analysis predicting that the serovar was responsible for 48% of all iNTS infections between 1966 and 2014 (Uche et al., 2017). Furthermore, between 2002 and 2008, MDR *S.* Typhimurium was responsible for an epidemic of bloodstream infections in Malawi. However, since 2008 the proportion of bloodstream infections caused by *S.* Typhimurium has significantly decreased (Feasey et al., 2015) potentially indicating that this serovar is becoming less important in modern day.

1.5.2 *S.* Typhimurium ST313

WGS showed that the majority of *S.* Typhimurium isolated from bloodstream infections in Africa belong to the sequence type (ST)313 (Kingsley et al., 2009). Subsequent phylogenetic analysis has identified two closely-related but genetically distinct lineages of ST313, differing from each other by 455 SNPs (Okoro et al., 2012). According to early spatial-temporal reconstructions of the ST313 phylogeny, the emergence of ST313 lineage 1 (L1) occurred in approximately 1960 in south west Africa, whereas ST313 lineage 2 (L2) occurred in approximately 1977 in Malawi, in close temporal correlation with the HIV epidemic (Okoro et al., 2012). The finished genome of ST313 L2 strain D23580, is widely used as the reference strain for ST313.

1.5.3 Antimicrobial resistance patterns in ST313 lineages

Both ST313 lineages are associated with resistance to multiple antibiotic agents, mostly encoded by various Tn*21*-like integrons inserted on the pSLT *Salmonella* virulence plasmid (Kingsley et al., 2009). The spread of these combined virulence-resistance-plasmids has increased rapidly in the past few decades and are believed to be important drivers of iNTS epidemics (Emond-Rheault et al., 2020). The characteristic AMR profile associated with ST313 lineages includes resistance to ampicillin, trimethoprim/sulfamethoxazole

(henceforth cotrimoxazole) and chloramphenicol (Feasey et al., 2014). Of these antimicrobials, chloramphenicol resistance is specifically found in ST313 L2, which is thought to have offered an evolutionary advantage, resulting in the clonal replacement of ST313 L1 by L2 in the early 2000s (Okoro et al., 2012). High levels of chloramphenicol resistance necessitated the use of alternative antimicrobials be used for the empirical management of iNTS in many countries in sub-Saharan Africa (Feasey et al., 2014). Consequently, ceftriaxone and fluoroquinolones are some of the few antimicrobial classes currently recommended for treatment of iNTS disease (Crump et al., 2015; Cuypers et al., 2018; Shane et al., 2017). Of concern, extensively drug resistant (XDR) isolates of ST313 have begun to emerge in Kenya (Kariuki et al., 2015; Oneko et al., 2015), Malawi (Feasey et al., 2014) and the Democratic Republic of Congo (DRC) (Van Puyvelde et al., 2019). The XDR profile is mediated by the presence of incompatibility group (Inc)HI2 plasmids which encode resistance to ciprofloxacin and extended spectrum beta-lactams (ESBLs) including ceftriaxone (Piccini and Montomoli, 2020).

1.5.4 Prophage repertoire

In addition to accessory genome elements relating to antimicrobial resistance, the majority of ST313 additionally carry two novel prophages named Blantyre Prophage 1 (BTP1) and Blantyre Prophage 5 (BTP5) (Kingsley et al., 2009; Okoro et al., 2012; Owen et al., 2017). Prophages encode many key *Salmonella* virulence factors (Brussow et al., 2004; Figueroa-Bossi et al., 2001; Ho and Slauch, 2001; Wagner and Waldor, 2002). For example, the common prophage Gifsy-2 carries the gene *sodC* and the SopEϕ prophage carries the guaninine nucleotide exchange factor (Figueroa-Bossi et al., 2001; Pelludat et al., 2003). A comprehensive investigation which characterised the prophage repertoire of ST313 revealed fascinating insights about BTP1 and posed interesting questions about the potential fitness costs and benefits of the prophage (Owen et al., 2017). For instance, BTP1 carries the *bstA* (*st313-td*) gene (*STMMW_03531*), which is associated with increased bacterial survival within murine macrophages (Herrero-Fresno et al., 2014). Additionally, the BstA protein of BTP1 is a novel abortive phage immunity system which mediates defence against attack by P22-like phages, and could improve fitness of ST313 in certain environmental niches (Owen et al., 2020).

It had originally been thought that the ST313 sequence type was geographically restricted to Africa. However, ST313 has also been isolated from Latin America, Asia and Europe (Almeida et al., 2017; Ashton et al., 2017; Jacob et al., 2019). For example, in 2017 it was discovered that 2.7% of *S.* Typhimurium isolated in England and Wales belong to ST313 and were associated with gastrointestinal infection (Ashton et al., 2017). Subsequent phylogenetic analysis identified a high level of genetic diversity amongst the UK-related ST313 (**Figure 1.5**). The two complete prophages (BTP1 and BTP5) characteristic of African ST313 were absent from the UK isolates, but were carried by the African L1 and L2 lineages of ST313 (Ashton et al., 2017).

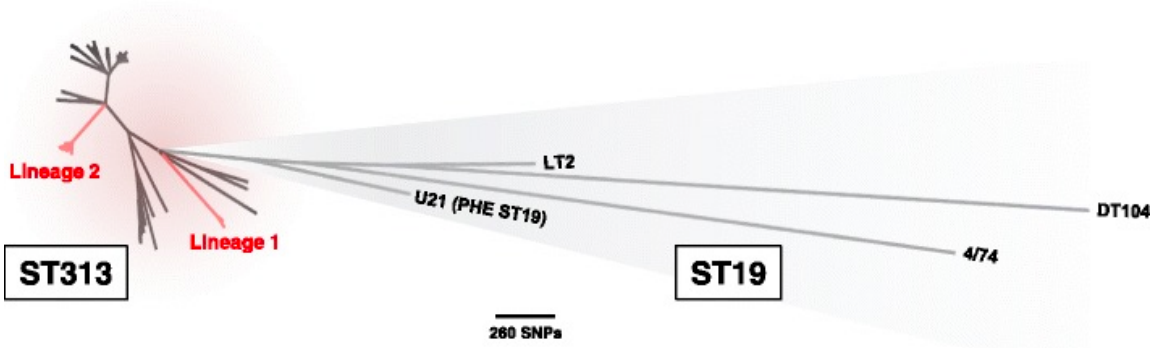


Figure 1.5 Phylogenetic relationship between ST313 lineages in the context of ST19 isolates

An unrooted maximum likelihood phylogeny is displayed, showing the clonality of ST313 L1 and ST313 L2 (red) compared with UK-isolated ST313 and ST19 representatives. **(Figure from Ashton *et al.*, 2017).**

1.5.5 Functional analysis of ST313

ST313 is closely related to ST19, the typical *S.* Typhimurium sequence type which causes self-limiting gastrointestinal infection globally. Pairwise comparisons between the ST19 reference strain, LT2 (McClelland et al., 2001) and the ST313 reference strain, D23580 has demonstrated 95% similarity in core genome composition (Canals et al., 2019). Genetic differences include 788 SNPs, three multi-nucleotide polymorphisms (MNPs), 65 indels, and an accumulation of ST313-specific pseudogenes (Canals et al., 2019). The majority of these pseudogene-forming SNPs affect metabolic, structural and virulence genes or are of unknown function (Okoro et al., 2015).

As previously discussed, genome degradation is a hallmark feature of systemic *Salmonella* which cause bloodstream infections and are adapting to occupy a restricted host range.

Accordingly, significant focus has been invested into understanding the functional relevance of pseudogene formation in ST313 (Aulicino et al., 2018; Bogomolnaya et al., 2013; Canals et al., 2019; Carden et al., 2017; Kingsley et al., 2009; Okoro et al., 2015; Singletary et al., 2016). As a pathogen adapts to a more restricted host range, degradation may occur in genes which are superfluous in the new environmental conditions. For example, a study in 2016 identified a loss of multicellular behaviour in ST313 L2 strain D23580 (Singletary et al., 2016). Specifically, two pseudogenes were investigated and shown to impair multicellular stress resistance, a phenotype associated with survival outside of the host. One of these mutations results in the inactivation of the *katE* gene which encodes a stationary phase catalase required to resist peroxide-mediated killing (Singletary et al., 2016). The second mutation affects the *bcsG* gene which encodes a cellulose biosynthetic enzyme required for the RDAR (red, dry and rough) colonial phenotype (Singletary et al., 2016). The loss of traits required for environmental stress-resistance supports mounting evidence for a human-human transmission of ST313 (Post et al., 2019).

Additional evidence of genome degradation, consistent with niche adaptation can be seen in genes associated with gut persistence in murine models (*ratB*), genes involved in the utilisation of certain carbon sources (*ttdA* and *melR*) and genes which encode effector proteins that are implicated in gastrointestinal pathogenesis (*sseI* and *pipD*). The SseI protein is particularly interesting as it inhibits the migration of dendritic cells into systemic sites by altering chemotaxis (Brink et al., 2018). However, studies within a streptomycin-treated murine infection model showed that pseudogenisation of *sseI* in ST313 L2 caused an increased ability of the bacteria to disseminate rapidly from the gut to the draining lymph nodes (Carden et al., 2017). Furthermore, the *sseI* gene was degraded or absent in multiple highly-adapted systemic *Salmonella* serovars, emphasising the importance of gene inactivation in promoting systemic spread (Feng et al., 2013; Klemm et al., 2016; Porwollik et al., 2005; Thomson et al., 2008).

The high incidence of iNTS disease in adults with HIV has led researchers to believe that ST313 is specifically adapting to a systemic lifestyle within the immunocompromised host (Feasey et al., 2012; Gordon, 2008). However, it wasn't until recently that phenotypic data was generated to support this hypothesis. The gene *macB* is a pseudogene in ST313 and forms part of the macAB-TolC tripartite efflux pump which is implicated in *Salmonella*

pathogenesis. Variants in *macAB* have lately been shown to affect replication in macrophages and influence fitness during colonisation of the murine gastrointestinal tract (Honeycutt et al., 2020). Importantly the infection outcome of *S.* Typhimurium strains carrying *macAB* variants were affected by a combination of the *Salmonella* genetic background and the host gene Nramp1, an important determinant of innate resistance to intracellular bacterial infection (Mastroeni, 2002). A summary of the pseudogenes identified in ST313 reference strains is shown in **Figure 1.6**.

STM	STMMW	Gene name	Description	D23580	U2
0157	01631	*yacH*	putative outer membrane protein		
0522	-	*allP*	allantoin transport protein		
0523	-	*allB*	allantoinase		
0834	08851	*ybiP*	putative integral membrane protein		
0942	09551	*macB*	putative ABC superfamily transport protein		
1014	10251		Probable regulation protein		
1023	10351		hypothetical on gifsy-2 prophage		
1051	10631	*sseI*	type III secretion effector protein (SPI-2)		
1092	-	*orfX*	putative cytoplasmic protein		
1093	-		putative cytoplasmic protein		
1094	11041	*pipD*	similar to dipeptidase A		
1228	12371		putative periplasmic protein		
1516	15161	*ydeE*	putative MFS family transport protein		
1548	15471		putative ribosyltransferase-isomerase		
1549	-		putative translation initiation inhibitor		
1550	-		putative cytoplasmic protein		
1551	-		putative cytoplasmic protein		
1551'	-		hypothetical protein		
1552	-		putative cytoplasmic protein		
1637	16321		putative inner membrane protein		
1862	-	*pagO*	PhoP activated gene		
1863	-		putative inner membrane protein		
1864	-		putative inner membrane protein		
1865	-		Putative DNA invertase		
1866	-		hypothetical		
1868	-	*mig-13*	phage tail assembly protein		
1868'	-		lytic enzyme		
1869'	-		hypothetical protein		
1870	-		RecE-like protein		
1896	18781		putative cytoplasmic protein		
1940	19221		putative cell wall associated hydrolase		
2238	22681		putative phage protein		
2514	25311	*ratB*	Secreted protein		
2589	26091		hypothetical in Gifsy-1 prophage		
2680	26941		putative cytoplasmic protein		
2932	28951	*ygbE*	putative inner membrane protein		
3012	29741		putative transcriptional regulator		
3075	30361		putative ABC-type cobalt transport system		
3355	33531		tartrate dehydratase		
3624	36131	*yhjU*	putative inner membrane protein		
3745	37341		putative cytoplasmic protein		
3768	37571		putative selenocysteine synthase		
4196	41451		putative cytoplasmic protein		
4286	42371	*lpxO*	putative dioxygenase		

Figure 1.6 Pseudogenes identified in ST313 reference strains

A list of pseudogenes identified in ST313 L2 reference strain D23580, with their putative functions. The conservation of these pseudogenes in the UK-isolated ST313 reference strain U2 is also shown, indicating differences in genome degradation. Specifically, grey indicates pseudogenes which are conserved in both strains and red indicates genes which are not degraded in the strain U2 and are likely to be functional **(Figure from Ashton *et al.,* 2017).**

1.6 An introduction to *S.* Panama

1.6.1 Clinical picture of *S.* Panama in humans

The high levels of iNTS disease caused by *S.* Typhimurium in sub-Saharan Africa has generated a significant research focus (Ao et al., 2015; Stanaway et al., 2019a). However, other *Salmonella* serovars associated with iNTS infections have received little attention. One of these neglected serovars is *S.* Panama*,* which has consistently been reported as one of the *Salmonella* serovars most frequently isolated over the past 70 years (Pulford et al., 2019). Although *S*. Panama can cause gastrointestinal infection in humans (Hendriksen et al., 2011), the serovar is more widely known for its ability to cause invasive disease and to colonise extraintestinal sites. For most salmonellae, extraintestinal colonisation refers to bloodstream infection (Ao et al., 2015). However, *S*. Panama can also invade specific body sites, causing atypical presentations, including throat infection, brain abscess, and Bartholin's abscess (Cummins and Atia, 1994; Kostiala et al., 1992; Varela and Ochoa, Aguilar, 1953) (summarized in **Figure 1.7**). These unexpected symptoms of *S*. Panama infection can impede diagnosis and delay treatment.

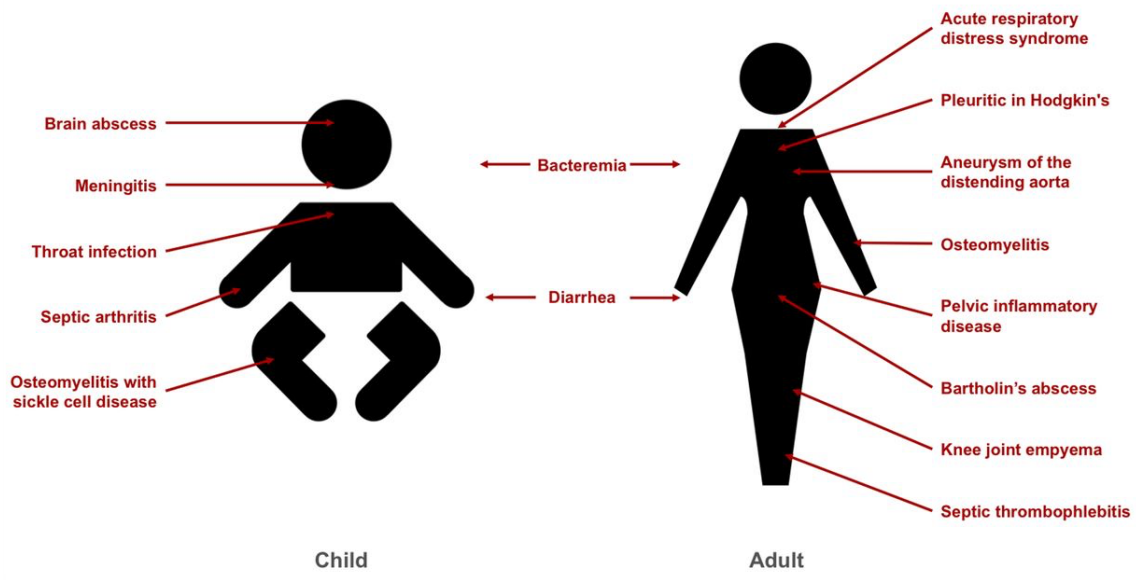


Figure 1.7 Overview of the clinical presentation caused by *S.* Panama in adults and children

S. Panama can cause a range of clinical syndromes (Barczi and Foldes, 1990; Borderon et al., 1981; Carneiro et al., 2018; Cordano and Virgilio, 1996; Cummins and Atia, 1994; Edel et al., 1978; Elenga et al., 2017; Gericke and Luchtrath, 1951; Heysell et al., 2008; Kostiala et al., 1992; Kostiala and Ranta, 1989; Lazarescu et al., 1971; Leeder, 1956; Modai et al., 1974; Opree, 1975; Salamon and Prag, 2001; Talon et al., 1985; van Cappelle et al., 1995; Varela and Ochoa, Aguilar, 1953) **(Figure from Pulford *et al.,* 2019).**

1.6.2 Global disease burden and epidemiology

The proportion of clinical cases caused by *S*. Panama is particularly high in French territories in America located in the Caribbean and South America (Guyomard-Rabenirina et al., 2018; Olive et al., 1996; Papa, 1976). Historically, *S*. Panama infections have accounted for 35% of *Salmonella* infections in Martinique, which were 11 times more likely to cause invasive disease than other serovars (Olive et al., 1996; Papa, 1976). Similarly, in French Guiana, *S*. Panama was the most frequent *Salmonella* serovar acquired by humans, accounting for 12.9% of all cases of *Salmonella* infection in 2011 (Gay et al., 2014). More recently, *S*. Panama was listed as the *Salmonella* serovar most frequently isolated from paediatric blood samples in Guadeloupe, contributing to one-third of all cases of *Salmonella* infection between 2010 and 2014; and univariate analysis showed *S*. Panama was associated with causing disease in children older than 6 months of age ($P = 0.002$) (Guyomard-Rabenirina et al., 2018).

S. Panama was also introduced into Europe, where the serovar spread through the pork industry and caused hospital outbreaks in the 1960s and 1970s. Over this period, there was a 3-fold increase in *S*. Panama cases in the United Kingdom, which caused a doubling of the number of salmonellosis cases (Lee, 1974). Subsequently, between 1969 and 1984, *S*. Panama was one of the top five serovars responsible for invasive disease in the United Kingdom (Wilkins and Roberts, 1988). Elsewhere in the European Union, *S*. Panama was reported among the top 10 most frequently isolated serovars during 2012, following 706 confirmed cases of *S*. Panama salmonellosis associated with outbreaks in Germany and Italy (European Food Safety Authority and European Centre for Disease Prevention and Control, 2012). Sporadic outbreaks of *S*. Panama salmonellosis also occurred in Switzerland (1972), Hungary (1979), Spain (1998), and the Netherlands (2008) (Ernst and Gurdan, 1973; Lantos et al., 1981; Noël et al., 2010; Soto et al., 2001). *S*. Panama maintained its ranking in the top 20 serovars associated with salmonellosis in the European Union until 2017 (European Food Safety Authority and European Centre for Disease Prevention and Control (EFSA and ECDC), 2019).

In Asia, *S*. Panama was the 11th most frequently isolated *Salmonella* serovar in humans between 2001 and 2007 (Hendriksen et al., 2011), with up to 70% of *S.* Panama isolates reported to be invasive compared to 12% of *S*. Enteritidis isolates (Huang et al., 2013). In Tokyo, Japan, *S*. Panama was the third most common *Salmonella* serovar between 1974 and

1979, accounting for 5% of cases of *Salmonella* infection, and was commonly isolated from asymptomatic people (Horiuchi et al., 1989). In Taiwan, where *S*. Panama causes 7% of the clinical cases of salmonellosis, *S*. Panama causes a higher rate of bacteraemia in children under 5 years of age than other serovars, such as *S*. Enteritidis (Tsai et al., 2007).

1.6.3 Antimicrobial resistance in *S.* Panama

There are conflicting reports in the literature relating to the AMR status of the *S*. Panama serovar. Studies in Italy and Brazil reported low levels of antibiotic resistance (Carneiro et al., 2018; Lamanna et al., 1977) and 91% of *S*. Panama isolates from Martinique were susceptible to beta-lactams (Olive et al., 1996). In Guadeloupe all *Salmonella* serovars demonstrated high overall susceptibility to antibiotics (Guyomard-Rabenirina et al., 2018). In contrast, other studies reported higher levels of resistance in *S*. Panama, particularly against tetracycline (e.g., 67%) and chloramphenicol (e.g., 67%) since the 1980s (Kempf and Pietzsch, 1977; Lee et al., 2008; Matsushita et al., 2001). In Asia, for example, *S*. Panama has been associated with high levels of AMR since 1980, when 58% of the *S*. Panama isolates from Tokyo were resistant to at least one antibiotic agent (Matsushita et al., 2001). The level of AMR in *S.* Panama appears to be on the rise. By the turn of the millennium, 83% of domestic and imported *S*. Panama isolates from cases in Tokyo were multidrug resistant. Similarly, in Taiwan, the serovar also exhibited resistance to multiple antibiotics, including cotrimoxazole (67%), ampicillin (56%), streptomycin (56%), kanamycin (56%), and gentamicin (45%) (Lee et al., 2008).

1.6.4 Accessory genome and virulence

A large proportion of *S*. Panama AMR has been associated with plasmid carriage ($P = 0.012$), class 1 integron presence, and transmissible drug resistance (R) factors (Huang et al., 2013; Lee et al., 2008; Rodríguez et al., 2006). Generally, plasmids play a key role in systemic *Salmonella* infection (Silva et al., 2017), but little is known about the plasmid complement of the *S*. Panama serovar. In the small number of available studies, it was reported that *S*. Panama does not commonly carry the large plasmids that have previously been associated with virulence in other *Salmonella* serovars (Carneiro et al., 2018). Rather, *S*. Panama strains carry a heterogeneous population of plasmids (Stanley et al., 1995). Prophages can also make significant contributions to *Salmonella* virulence (Graham et al., 2018; Wahl et al., 2019), but only one study has identified the presence of prophages in *S*. Panama (Yao et al., 2016). The *Salmonella* RE-2010 prophage was identified in the genome of *S*. Panama ATCC 7378. The

prophage (also known as ElPhiS) has also been found in *S*. Enteritidis, where it has been associated with specific phylogenetic clusters (Feasey et al., 2016; Graham et al., 2018; Hanna et al., 2012). The paucity of authoritative literature describing collections of *S.* Panama isolates, coupled with the importance of *S*. Panama for global public health made it imperative that a concerted comparative genomic analysis was conducted.

1.6.5 Evolutionary history and virulence

The study of evolutionary history could explain why *S*. Panama is associated with invasive disease. Phylogenetically, *S*. Panama belongs to *S. enterica* clade B, which is associated with increased levels of clinically invasive disease (den Bakker et al., 2011). The evolutionary history of *S*. Panama was initially studied by Selander and colleagues who used multilocus enzyme electrophoresis to assess the relationships among *Salmonella* serovars that cause invasive disease (Selander et al., 1990). It was proposed that *S*. Panama evolved from the same ancestors that gave rise to *S.* Paratyphi, *S.* Sendai (which causes enteric fever), and *S.* Miami (Selander et al., 1990). However, no large-scale phylogenetic study had yet been conducted on *S*. Panama, and only one complete *S*. Panama genome sequence is available (Yao et al., 2016).

1.7 Aims of thesis

1.7.1 Background summary

NTS is a globally important pathogen, typically associated with foodborne gastrointestinal infection (Kirk et al., 2015). However, NTS can also invade normally sterile sites in humans resulting in invasive disease such as bacteraemia or meningitis, often in immunosuppressed individuals (Feasey et al., 2012; Gordon, 2008; Stanaway et al., 2019a). This systemic infection, known as iNTS disease is typically associated with certain *Salmonella* serovars (Jones et al., 2008). Thus, the clinical presentation of NTS disease depends on a combination host immune factors and bacterial features that are specific to individual *Salmonella* pathovariants (de Jong et al., 2012; Gilchrist and MacLennan, 2019; Lokken et al., 2016). It was initially believed that the evolutionary trajectory from NTS to iNTS would be driven primarily by the acquisition of beneficial virulence factors, prophages and plasmids. However, with the increasing availability of WGS and subsequent genomic exploration, a much more complex picture has emerged (Loman and Pallen, 2015; Pallen and Wren, 2007). Specifically, there is a paradigm of bacterial entrance into a more restricted host range coupled with a reduced metabolic flexibility and the loss of factors that have become unnecessary within the pathogen's new niche (Kingsley and Bäumler, 2000; Okoro et al., 2015). A similar pattern is observed in the emergence of other important human pathogens from non-pathogenic bacterial species such as *Yersinia* and *Shigella* (McNally et al., 2016; Merhej et al., 2013; Reuter et al., 2014; The et al., 2016).

1.7.2 Thesis aim

In the WGS-informed era of pathogen surveillance and scientific research, unravelling the genomic epidemiology and evolutionary history of iNTS serovars will significantly advance knowledge on key invasive *Salmonella* pathovariants whilst simultaneously providing vital understanding for the prevention of future outbreaks and endemics. **In this thesis I aimed to determine the temporal, geographical and evolutionary trends in nontyphoidal *Salmonella* serovars, in order to understand the genomic signatures acquired on the trajectory towards extraintestinal adaptation.**

1.7.3 Overview of work

In Chapter 2, I investigated the diversity of 133 NTS isolates sampled from venomous snakes and non-venomous reptiles. By studying a unique dataset, I have revealed substantial

diversity among the serovars carried by reptiles. This broad scope of *Salmonella* serovars provided a good sample population for studying the divergence and radiation of almost the entire *Salmonella* genus into subspecies, clades, serovars and lineages. which facilitated the study of *S. enterica* species level evolution and ecology. I also conducted a genotypic and phenotypic investigation into metabolic differences between phylogenetic clades, highlighting signatures of niche adaptation in a broad scope of NTS serovars. This broad snapshot of *Salmonella* provided the context to then look in more detail at two epidemiologically important serovars associated with invasive infection in humans.

In Chapter 3, I characterised the stepwise evolution of *S.* Typhimurium ST313 associated with bloodstream infection of humans in Africa. Using WGS data from 680 human *S.* Typhimurium bloodstream isolates collected across sub-Saharan Africa, I have identified the trajectory of crucial loss-of-function events along the evolutionary pathway of iNTS. Furthermore, I have uncovered a pan-susceptible ST313 lineage (ST313) which emerged in Malawi in 2016 and has an elevated invasiveness index compared to previously described ST313 lineages. My work here provides an expanded insight into the evolutionary dynamics of *S.* Typhimurium ST313 and highlights epidemiological shifts in the circulation of ST313 lineages in Malawi. In order to contrast this work, I looked at a less-well studied serovar responsible for extraintestinal infection globally.

In Chapter 4, I determined the genomic epidemiology of 489 *S.* Panama from 27 countries collected over 26 years during routine public health surveillance. I identified six geographically-localised *S.* Panama clades and revealed regional trends in AMR. Further phylodynamic investigation revealed spatial-temporal evolution of the serovar, providing a vital baseline of understanding for *S.* Panama infection in the future.

Collectively, this thesis has utilised WGS-approaches to characterise a broad range of NTS serovars and complemented this investigation by more detailed studies of two serovars; one well-characterised and one understudied, both of which have caused outbreaks, endemics and epidemics of extraintestinal infections. By exploring the genomic epidemiology of key *Salmonella* pathovariants associated with invasive disease globally, I have resolved gaps in the literature by describing a succession of events and linking them with functional genomic information that contribute to the emergence of iNTS serovars and pathovariants.

Chapter 2

The diversity, evolution and ecology of *Salmonella* in venomous snakes

2.1 Acknowledgement of the specific contribution of collaborators

Much of the content of this chapter was published in the research article "The diversity, evolution and ecology of *Salmonella* in venomous snakes" (Pulford *et al.* 2019) in PLOS Neglected Tropical Diseases. Permission to include the publication in this PhD thesis was obtained from all co-authors. Specifically, I acknowledge the following contribution of collaborators to the experimental work described in this chapter. Unless specified below, all work was completed by myself.

Nicolas Wenner University of Liverpool, UK	Assisted in methodology design and supervision.
Martha L. Redway University of Liverpool, UK	Assisted with the isolation of *Salmonella* from faecal samples collected from venomous reptiles housed at Liverpool School of Tropical Medicine* and phenotypic characterisation of AMR profiles.
Ella V. Rodwell University of Liverpool, UK	Assisted with phenotypic characterisation of carbon source utilisation profiles.
Hermione J. Webster University of Liverpool, UK	Assisted with phenotypic characterisation of carbon source utilisation profiles.
Roberta Escudero University of Liverpool, UK	Assisted with the isolation of *Salmonella* from faecal samples collected from venomous reptiles housed at Liverpool School of Tropical Medicine*.
Carsten Kröger University of Liverpool, UK	Assisted in methodological design and supervision of Roberta Escudero.
Rocío Canals University of Liverpool, UK	Assisted with phenotypic characterisation of carbon utilisation profiles involving anaerobic metabolism.
Will Rowe University of Liverpool, UK	Assisted in producing bioinformatics analysis pipeline.
Javier Lopez Chester Zoo, UK	Collection of faecal samples from non-venomous retiles housed in Chester Zoo*.
Neil Hall Earlham Institute, UK	Sequencing of isolates as part of the 10,000 *Salmonella* genomes project.
Paul D. Rowley Liverpool School of Tropical Medicine, UK	Collection of faecal samples from venomous reptiles housed at Liverpool School of Tropical Medicine*.
Dorina Timofte University of Liverpool, UK	Provided access to isolate collection at University of Liverpool Leahurst Campus
Robert A. Harrison Liverpool School of Tropical Medicine, UK	Supervision and conceptualisation of project and provided access to samples at LSTM
Kate S. Baker University of Liverpool, UK	Supervision, conceptualization and methodology design.
Jay C.D. Hinton University of Liverpool, UK	Supervision, conceptualization and methodology design.

*Please note that faecal sample collection and *Salmonella* isolation was largely performed prior to beginning my PhD project and was not done as part of this thesis. However, the methods are included for clarity. **This only applies to the work described in Section 2.3.2 and where explicitly mentioned in Section 2.4.1.** Specifically, sample collection was either performed by myself or as listed above.

2.2 Introduction

The *Salmonella* genus contains two species; *S. bongori* and *S. enterica. S. enterica* is further divided into six subspecies; *enterica* (I), *salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb), *houtanae* (IV) and *indica* (VI) (Brenner et al., 2000). The subspecies are classified into approximately 2,600 serovars which are ecologically, phenotypically and genetically diverse (Winfield and Groisman, 2004). Serovars which belong to *S. enterica* subspecies *enterica* cluster phylogenetically into two predominant clades (A and B) (den Bakker et al., 2011; Falush et al., 2006; Parsons et al., 2011). Metabolic properties such as carbon utilisation are often serovar-specific. The ability of *Salmonella* to grow in a wide range of conditions reflects the adaptation of the bacteria to survival in the environment or in different hosts, as demonstrated by a recent study focused on genome-scale metabolic models for 410 *Salmonella* isolates spanning 64 serovars in 530 different growth conditions (Seif et al., 2018).

At the genus level, *Salmonella* has a broad host-range whilst individual serovars differ in host-specificity (Bäumler and Fang, 2013). The majority of *Salmonella* infections in humans (99%) are caused by a small number of serovars belonging to the *S. enterica* subspecies (Eng et al., 2015). Serovars which belong to non-*enterica* subspecies are associated with carriage in ectothermic animals such as reptiles and amphibians, but are rarely found in humans (Bäumler and Fang, 2013; Clancy et al., 2016; Nair et al., 2014; Schroter et al., 2004). Carriage rates of non-*enterica* serovars in reptiles can be high. A study focused on snakes in a pet shop found that 81% of animals carried *S. diarizonae* (Schroter et al., 2004). Previous studies demonstrating the diverse range of *Salmonella* serovars that colonise various reptilian species in different countries are summarised in **Table 2.1**.

Reptiles represent a significant reservoir for serovars of *Salmonella* that are associated with human disease (Guyomard-Rabenirina et al., 2019). Thus, *Salmonella* isolated from reptiles represents an ideal sample population to study serovar-specific adaptation and overall population structure. Over 60% of captive-bred reptiles between 1995 and 2006 in Denmark were reported to carry *S. enterica* subspecies *enterica* serovars (Pedersen et al., 2009). About 6% of human salmonellosis cases were contracted from reptiles in the USA (Mermin et al., 2004), and in South West England, 27.4% of *Salmonella* cases in children under five years old were linked to reptile exposure (Murphy and Oshin, 2015). The latter study demonstrated

that reptile-derived salmonellosis was more likely to cause bloodstream infection in humans than non-reptile-derived *Salmonella* (Murphy and Oshin, 2015). Reptile-associated *Salmonella* is therefore considered to be a global threat to public health (Whiley et al., 2017).

The majority of reptile-associated salmonellosis cases reported in humans are caused by *Salmonella* from non-venomous reptiles (Whiley et al., 2017), probably because these animals are frequently kept as pets. Therefore, non-venomous reptiles have been the focus of numerous studies whilst the prevalence and diversity of *Salmonella* in venomous snakes has not been investigated. The recent inclusion of snakebite as a neglected tropical disease (Chippaux, 2017) demonstrates that these reptiles frequently interact with humans in tropical and sub-tropical countries. The proximity of venomous snakes to humans may lead to contaminated faecal matter being shed on the surfaces and in water sources used for human homes and to irrigate salad crops (Harrison et al., 2009; Ogunfowokan, 2012; Warrell et al., 1976). Research to improve snakebite treatment at the Liverpool School of Tropical Medicine (LSTM) led to the creation of the most extensive collection of venomous snakes in the UK, containing 195 venomous snakes. The LSTM herpetarium houses venomous snakes from a diverse range of species and geographical origins, representing an ideal source of samples to assess *Salmonella* in this under-studied group of reptiles.

The aims of this study were three-fold. Firstly, to determine the period prevalence of *Salmonella* in a collection of captive venomous snakes and investigate whether this group of reptiles had the potential to act as reservoirs for human salmonellosis. Secondly, to assess the serological and phylogenetic diversity of *Salmonella* amongst reptiles. Thirdly, to use the diversity of reptile-associated *Salmonella* to determine clade-specific differences that could reflect adaptation to survival in different niches. Here, I present the first whole genome-based comparative study of the *Salmonella* bacteria that colonise venomous and non-venomous reptiles.

Table 2.1 The carriage of *Salmonella* subspecies by non-venomous reptiles

Country of study	Reptile information	Proportion of reptiles carrying *Salmonella*	Serovar Composition	Citation
Poland	Snakes Lizards Chelonians	122 of 374 (32.6%)	59% *enterica* 16% *salamae* or *houtanae* 3.5% *diarizonae* or *indica*	(Piasecki et al., 2014)
Croatia	Snakes Lizards Chelonians	26 of 200 (13.0%)	34.6% *enterica* 23.1% *houtanae* 23.1% *arizonae* 15.4% *diarizonae* 2.8% *salamae*	(Lukac et al., 2015)
Japan	Snakes Lizards Turtles	83 of 112 (74.1%)	62.5% *enterica* 37.5% *salamae, arizonae, diarizonae* or *houtanae*	(Nakadai et al., 2005)
Germany and Austria	Reptiles	86 of 159 (54.1%)	52.8% *enterica* 34.5% *diarizonae* 3.4% *salamae* 6.9% *arizonae* 2.3% *houtanae*	(Geue and Loschner, 2002)
Brazil	Pet Snakes Lizards Chelonians	38 of 97 (39.1%)	44.7% *enterica* 10.5% *salamae* 5.2% *arizonae* 21% *diarizonae* 18.5% *houtanae*	(Sá and Solari, 2001)
Denmark	Captive reptiles	Not discussed	65% *enterica* 12% *diarizonae* 11% *houtanae* 6% *salamae* 6% *arizonae*	(Pedersen et al., 2009)
Germany	Captive reptiles	13 of 16 (81.3%)	100% *diarizonae*	(Schroter et al., 2004)

2.3 Methods

2.3.1 Source of *Salmonella* isolates

The *Salmonella* isolates were derived from faecal samples from two collections of reptiles. One hundred and six faecal samples were collected from venomous snakes at LSTM, with an emphasis on snakes originating from Africa (**Table S2.1**), and investigated for the presence of *Salmonella.* All venomous snakes were housed in individual enclosures and fed with frozen mice provided by a single supplier. Sixty-nine of the samples (71%) were sourced from wild-caught snakes originating from: Togo, Nigeria, Cameroon, Egypt, Tanzania, Kenya, South Africa, and Uganda. A further 28 *Salmonella* isolates (29%) came from venomous snakes bred in captivity. The LSTM herpetarium is a UK Home Office licensed and inspected animal holding facility. A second collection of 27 *Salmonella* isolates from non-venomous reptiles and 1 *Salmonella* isolate from a venomous reptile were sourced from the veterinary diagnostics laboratory based at the University of Liverpool's Leahurst campus (reptilian species described in **Table S2.1**). These isolates were collected from specimens submitted as part of *Salmonella* surveillance for import/export, in addition to veterinary faecal samples and tissues from post-mortem investigations. The provenance of the isolates is described in **Table S2.1**. The majority of the non-venomous reptiles were sourced from a zoological collection, however two animals were privately owned and three were sourced from the Royal Society for the Prevention of Cruelty to Animals (RSPCA). The LSTM isolates are henceforth referred to as venomous snake isolates and the Leahurst isolates are referred to as non-venomous reptile isolates unless otherwise stated.

2.3.2 Isolation of *Salmonella*

All media were prepared and used in accordance with the manufacturer's guidelines unless otherwise stated. *Salmonella* bacteria were isolated using a modified version of the protocol described in the national Standard Operating Procedure for detection of *Salmonella* issued by Public Health England (PHE)(Public Health England, 2014).

Faecal droppings were collected from reptiles and stored in 15 mL plastic centrifuge tubes at 4˚C. Two different methods were used for the enrichment of *Salmonella* from faecal samples due to reagent availability at the time of isolation. **Table S2.1** provides information on isolate specific methods. In enrichment method 1, faecal samples were added to 10 mL of buffered peptone water (Fluka Analytical, UK, 08105-500G-F) and incubated overnight at 37˚C with

shaking at 220 rpm. Following overnight incubation, 100 µL of the faeces mixture was added to 10 mL of Selenite Broth (19 g/L selenite broth base, Merck, UK, 70153-500G and 4 g/L sodium hydrogen selenite, Merck, UK, 1.06340-50G) and incubated overnight at 37°C with shaking at 220 rpm. In enrichment method 2, faecal samples were added to 10 mL of Buffered Peptone Water (Fluka Analytical, 08105-500G-F) supplemented with 10 µg/mL Novobiocin (Merck, N1628), and incubated overnight at 37˚C with shaking at 220 rpm. Following overnight incubation, 100 µL of the faeces mixture was added to 10 mL Rappaport-Vassilliadis Medium (Lab M, UK, LAB086) and incubated for 24 hours at 42˚C with shaking at 220 rpm.

Following enrichment by method 1 or 2, 10 µL of overnight broth was spread onto Xylose Lysine Deoxycholate (XLD) (Oxoid, UK, CM0469) agar plates which were incubated overnight at 37˚C. Putative *Salmonella* colonies were selected by black appearance on XLD plates and confirmed by pink and white colony formation on Brilliant Green Agar (Merck, 70134-500G) supplemented with 0.35 g/L mandelic acid (Merck, M2101) and 1 g/L sodium sulfacetamide (Merck, S8647).

To identify *S. enterica* species, colony PCR of the *Salmonella*-specific *ttr* locus, which is required for tetrathionate respiration (Malorny et al., 2004), was performed. PCR reagents included MyTaq Red Mix 1x (Bioline, UK, BIO-25043), *ttr-4* reverse primer (5'-AGCTCAGACCAAAAGTGACCATC-3') and *ttr-6* forward primer (5'-CTCACCAGGAGATTACAACATGG-3') on colonies suspected to be *Salmonella*. PCR reaction conditions were as follows: 95˚C 2 min, 35 x (95˚C 15 s, 60˚C 30 s, 72˚C 10 s), 72˚C 5 min. PCR products were visualised using agarose gel (3.5%) (Bioline, BIO-41025) electrophoresis in tris-acetate-EDTA (TAE) buffer. Midori Green DNA stain (3 µL/100 mL) (Nippon Genetics, Germany, MG 04) was used to visualise DNA bands under Ultra Violet light. Throughout the isolation procedure, *S.* Typhimurium strain LT2 (Beltran et al., 1988) was used as a positive control, and *E. coli* MG1655 (Blattner et al., 1997) was used as a negative control (**Table S2.1**).

2.3.3 Whole-genome sequencing of *Salmonella* from venomous and non-venomous reptiles

All non-venomous reptile isolates (*n* = 27), one venomous reptile isolate, and the majority of venomous snake isolates (*n* = 87/97) were sent for WGS. Isolates were sent for WGS during

or before December 2016 to accommodate project timescales. Because ten venomous snake isolates were collected in January 2017, they were unable to be sent for sequencing. Isolates were sent to either MicrobesNG, UK (*n* = 55) or the Earlham Institute, UK (*n* =60) for WGS on the Illumina HiSeq platform (Illumina, California, USA). Isolates which were sequenced by MicrobesNG were prepared for sequencing in accordance with the company's preparation protocol for single colony-derived bacterial cultures (https://microbesng.com/documents/3/MicrobesNG_preparing_your_bead_tubes.pdf).

Isolates which were sequenced by the Earlham Institute were prepared by inoculating a single colony of *Salmonella* into a FluidX® 2D Sequencing Tube (FluidX Ltd, UK) containing 100 µL of Lysogeny Broth (LB, Lennox) and incubating overnight at 37˚C, with shaking at 220 rpm. LB was made using 10 g/L Bacto tryptone (BD Biosciences, UK, 211705), 5 g/L bacto yeast extract (BD, 212750) and 5 g/L sodium chloride (Merck, S3014-1kg). Following overnight growth, the FluidX® 2D Tubes were placed in a 95˚C oven for 20 minutes to heat-kill the isolates.

DNA extractions and Illumina library preparations were conducted using automated robots at MicrobesNG and the Earlham Institute. At the Earlham Institute, the Illumina Nextera XT DNA Library Prep Kit (Illumina, FC-131-1096) was used for library preparation. High throughput sequencing was performed using an Illumina HiSeq 4000 sequencing machine to generate 150 bp paired-end reads. Sequencing was multiplexed with 768 unique barcode combinations per sequencing lane. The insert size was approximately 180 bp, and the median depth of coverage was 30x.

At MicrobesNG, genomic DNA libraries were prepared using the Nextera XT Library Prep Kit (Illumina, FC-131-1096) with two nanograms of DNA used as input and double the elongation time described by the manufacturer. Libraries were sequenced on the Illumina HiSeq 2500 using a 250 bp protocol.

2.3.4 Obtaining contextual reference sequences

The *Salmonella in silico* typing resource (SISTR) v1.0.2 was used for serovar prediction (Yoshida et al., 2016). Enterobase (Alikhan et al., 2018) was used to assign an MLST to each isolate, based on sequence conservation of seven housekeeping genes (Alikhan et al., 2018).

Where available, reference isolates representing previously sequenced *Salmonella* isolates for all subspecies and serovars identified were included in the analysis. Reference sequence assemblies were downloaded from the NCBI. Accession numbers are available in **Table S2.2.**

2.3.5 Quality control checks

FastQC v0.11.5 (www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC v1.0 (http://multiqc.info) were used to assess sequence read quality against the following criteria; passed basic quality statistics, per base sequence quality, per base N content, adapter content and an average GC content of between 47 % and 57 %. Only high-quality reads were used in downstream analysis. Kraken v0.10.5-beta (Wood and Salzberg, 2014) was run to ensure reads were free from contamination using the MiniKraken 8gb database and a *Salmonella* abundance cut-off of 70%. Trimmomatic v0.36 (Bolger et al., 2014) was then used on the paired-end reads to trim low-quality regions using a sliding widow of 4:200. ILLUMINACLIP was used to remove adapter sequences.

2.3.6 Phylogenetics- core gene alignment tree

Genomes were assembled using SPADES v3.90 (Bankevich et al., 2012). The Quality Assessment Tool for genome Assemblies (QUAST) v4.6.3 (Gurevich et al., 2013) was used to assess the quality of assemblies, the results of which can be found in **Table S2.3**. Assemblies which comprised of greater than 500 contigs failed quality control (QC) as they were deemed too fragmented for downstream analysis. Additional criteria to evaluate assembly quality to standards consistent with EnteroBase (Alikhan et al., 2018) was also assessed. Specifically, samples were excluded if they had an N50 less than 20 kb or if they had a total number of bases below 4Million base pairs (Mbp) or above 5.8Mbp. All assemblies which passed QC were annotated using Prokka v1.12 (Seemann, 2014). Roary v3.11.0 (Page et al., 2015) was used to generate a core gene alignment. SNP-sites v2.3.3 (Page et al., 2016) was used to extract SNPs. A maximum likelihood tree was built from the core gene SNP alignment of all isolates using RAxML-NG v0.4.1 BETA (Stamatakis, 2014) with the GTR model and gamma distribution for site-specific variation and 100 bootstrap replicates to assess support. The tree was rooted using the *Salmonella* species *S. bongori.* Interactive Tree Of Life v4.2 (Letunic and Bork, 2016) was used for tree visualisation.

Confirmation that there was no bias in phylogenetic signal between the two different sequencing platforms was achieved by mapping metadata onto the phylogenetic tree using iTOL (Letunic and Bork, 2016) and visually assessing clustering. **Table S2.1** contains details of the sequencing facility used for each genome. Monophyletic clustering of isolates was used to assign subspecies to newly sequenced *Salmonella* isolates from venomous and non-venomous reptilian hosts. The level of association between venom status and phylogenetic clade was determined using odds ratios and $\chi 2$ statistics using the OpenEpi website (http://www.openepi.com).

2.3.7 Identification of clade-specific genomic regions

Genes involved in the utilisation of individual carbon sources were identified using KEGG (Kanehisa et al., 2014) and relevant literature (**Table 2.2**). Genes that encoded proteins involved in the uptake of carbon sources were prioritised. Sequences were downloaded using the online tool SalComMac (Srikumar et al., 2015), which allows the download of fasta sequences of the genes from *S.* Typhimurium ST19 strain 4/74. In the case of *lac* genes, the sequences were taken from the *E. coli* reference sequence MG1655. The sequences can be found in **Data S2.1**.

The software tool MEGABLAST v2.2.17 (Morgulis et al., 2008) was used to perform a BLAST (Basic Local Alignment Search Tool) search of genes in the reptile-derived genomes against a custom-made database of genes diagnostic for *Salmonella* Pathogenicity Islands and genes involved in carbon utilisation. To confirm all MEGABLAST results, the short reads were mapped against each gene using BWA v0.7.10 (Li and Durbin, 2009) and SAMtools v0.1.19 (H. Li et al., 2009). The resulting bam files were manually assessed for gene presence and absence using the Integrative Genomics Viewer v2.4.15 (Robinson et al., 2011). The results were plotted against the maximum likelihood phylogeny using Interactive Tree Of Life (iTOL) v4.2 (Letunic and Bork, 2016).

Table 2.2 Carbon source utilisation gene information

Carbon Source	Gene	Gene product	LT2 Locus ID	SL1344 Locus ID	KEGG Pathway	Citation demonstrating gene activity in *Salmonella*
Glucose	*crr*	PTS system, glucose-specific IIA component	STM2433	SL1344_2396	sey00010	(Steeb et al., 2013)
	ptsG	PTS system, glucose-specific IIBC component	STM1203	SL1344_1140	sey00010	(Steeb et al., 2013)
Tricarballylic Acid	*tcuA*	FAD-Dependent tricarballylate dehydrogenase	STM0691	SL1344_0673	NA	(Lewis et al., 2004; Lewis and Escalante-Semerena, 2006)
	tcuB	Citrate/tricarballylate utilization protein TcuB	STM0690	SL1344_0672	NA	(Lewis et al., 2004)
	tcuC	Citrate-proton symporter TcuC	STM0689	SL1344_0671	NA	(Lewis et al., 2004)
	tcuR	LysR-family transcriptional regulator TcuR	STM0692	SL1344_0674	NA	(Lewis et al., 2004)
Maltose	*malE*	Periplasmic maltose-binding protein	STM4229	SL1344_4166	sey02010	(Daus et al., 2007)
	malF	Maltose transport inner membrane protein	STM4228	SL1344_4164	sey02010	(Daus et al., 2007)
	malG	Maltose transport inner membrane protein	STM4227	SL1344_4163	sey02010	(Daus et al., 2007)
	malK	Maltose transport ATP-binding protein	STM4230	SL1344_4167	sey02010	(Daus et al., 2007)
Acetate	*ackA*	Acetate kinase	STM2337	SL1344_2306	sey00620	(Fox and Roseman, 1986)
	pta	Phosphate acetyltransferase	STM2338	SL1344_2307	sey00620	(Starai et al., 2005)
Fructose	*fruA*	PTS system, fructose-specific IIBC component	STM2204	SL1344_2181	sey00051	(Feldheim et al., 1990)
	fruB	PTS system, fructose-specific IIA/FPR component	STM2206	SL1344_2183	sey00051	(Feldheim et al., 1990)
Melibiose	*melA*	Alpha-galactosidase	STM4298	SL1344_4235	sey00052	(Canals et al., 2019)
	melB	Melibiose carrier protein	STM4299	SL1344_4236	NA	(Canals et al., 2019)
	melR	Melibiose operon regulatory protein	STM4297	SL1344_4234	NA	(Canals et al., 2019)
Glucosamine	*nagA*	N-acetylglucosamine-6-phosphate deacetylase	STM0683	SL1344_0665	sey00520	(Steeb et al., 2013)
	nagB	Glucosamine-6-phosphate deaminase	STM0684	SL1344_0666	sey00520	(Steeb et al., 2013)
	nagE	N-acetylglucosamine-specific IIABC component	STM0685	SL1344_0667	sey00520	(Steeb et al., 2013)
	ycfX	N-acetyl-D-glucosamine kinase	STM1220	SL1344_1157	stm00520	(Steeb et al., 2013)
Glycerol	*gldA*	Glycerol dehydrogenase	STM4108	SL1344_4058	sey00561	(Steeb et al., 2013)
	glpK	Glycerol kinase	STM4086	SL1344_4035	sey00561	(Steeb et al., 2013)
Lactose	*lacY*	MFS transporter, OHS family, lactose permease	NA	NA	NA	(Leonard et al., 2015)
	lacZ	Beta-galactosidase	NA	NA	eco00052	(Leonard et al., 2015)
Tartaric acid	*ttdA*	Tartrate dehydratase	STM3355	SL1344_3327	sey00630	(Malorny et al., 2004)
	ttdB	Tartrate dehydratase	STM3354	SL1344_3326	sey00630	(Malorny et al., 2004)
Dulcitol	*gatA*	PTS system, galactitol-specific IIA component	STM3258	SL1344_3231	sey00052	(Nolle et al., 2017)
	gatB	PTS system, galactitol-specific IIB component	STM3259	SL1344_3232	sey00052	(Nolle et al., 2017)
	gatC	PTS system, galactitol-specific IIC component	STM3260	SL1344_3233	sey00052	(Nolle et al., 2017)
	gatD	Galactitol-1-phosphate dehydrogenase	STM3261	SL1344_3234	sey00052	(Nolle et al., 2017)
	gatR	Galactitol utilization operon repressor	STM3262	SL1344_3235	NA	(Nolle et al., 2017)
	gatY	Tagatose-bisphosphate aldolase	STM3253	SL1344_3226	sey00052	(Nolle et al., 2017)
Myo-inositol	*iolA*	Myo-inositol metabolism protein	STM4421	SL1344_4354	NA	(Kröger and Fuchs, 2009)
	iolB	Myo-inositol metabolism protein	STM4420	SL1344_4353	NA	(Kröger and Fuchs, 2009)
	iolR	Repressor of myo-inositol metabolism	STM4417	SL1344_4350	NA	(Kröger and Fuchs, 2009)
	iolT1	Myo-inositol transport protein IolT1	STM4418	SL1344_4351	NA	(Kröger and Fuchs, 2009)
	iolT2	Myo-inositol transport protein IolT2	STM4419	SL1344_4352	NA	(Kröger and Fuchs, 2009)
Allantoin	*allB*	Allantoinase	STM0523	SL1344_0516	sey00230	(den Bakker et al., 2011; Okoro et al., 2015)
	allP	Hypothetical allantoin permease	STM0522	SL1344_0515	NA	(den Bakker et al., 2011; Okoro et al., 2015)
	gcl	Glyoxylate carboligase	STM0517	SL1344_0510	NA	(den Bakker et al., 2011; Okoro et al., 2015)

2.3.8 Carbon source utilisation

Differential carbon source utilisation of 39 reptile-derived *Salmonella* isolates from *S. diarizonae, S. enterica* clade A and *S. enterica* clade B was assessed experimentally. Sub-sampling was necessary to accommodate experimental restrictions on the number of isolates that could be tested. The sub-sampling encompassed a range of serovars, location origins and reptile species. Filter-sterilised carbon sugar solutions were added into M9 (Merck, M6030-1kg) agar at concentrations detailed in **Table 2.3**. Isolated colonies were transferred from LB agar plates onto M9 carbon source plates using a sterile 48-pronged replica plate stamp and incubated at 37°C under aerobic conditions. An LB control plate was used to validate successful bacterial transfer and all experiments were performed in duplicate. If no growth was seen under aerobic conditions for a particular carbon source, the procedure was repeated under anaerobic conditions (approx. 0.35% oxygen) with 20 mM Trimethylamine *N*-oxide dihydrate (TMAO) (Merck, 92277) as a terminal electron acceptor. Anaerobic conditions were achieved by incubating plates in an anaerobic jar with 3x AnaeroGen 2.5 L sachets (Thermo Scientific, UK, AN0025A) to generate anaerobic gas. Oxygen levels were measured using SP-PSt7-NAU Sensor Spots and the Microx 4 oxygen detection system (PreSens, Regensburg, Germany). *Salmonella* growth was determined at 18, 90 and 162 hours in aerobic growth conditions and at 162 hours in anaerobic growth conditions. A subset of growth positive isolates were assessed for single colony formation to validate the results of the replica plating.

Table 2.3 Carbon source specific growth requirements

Carbon source	Final concentration in solid media	Aerobic/Anaerobic incubation	Product Details	References
Sodium acetate	0.60%	Aerobic	Merck, S2889-1KG	(Nichols et al., 2011)
Allantoin	0.10%	Anaerobic	Merck, 5670	(Feasey et al., 2016)
D-fructose	0.40%	Aerobic	Merck, F0127-100G	(Feasey et al., 2016)
D-glucosamine	0.40%	Aerobic	Merck, A15532	(Feasey et al., 2016)
D-melibose	0.40%	Aerobic	Merck, M5500-5G	(Feasey et al., 2016)
Dulcitol	0.40%	Aerobic	Merck, D0256-100G	(Feasey et al., 2016)
Glucose	0.40%	Aerobic	Merck, G7021	(Nichols et al., 2011)
Glycerol	0.40%	Aerobic	F, G/0650/17	(Nichols et al., 2011)
Maltose	0.40%	Aerobic	Merck, 1.0591	(Nichols et al., 2011)
Myo-inositol	0.40%	Aerobic	Merck, I7508-100G	(Gutnick et al., 1969)
L-tartaric acid	0.40%	Anaerobic	Merck, T109-500G-A	(Gutnick et al., 1969)
Tricarballylic acid	0.40%	Aerobic	Merck, T53503	(Gutnick et al., 1969)
Lactose	0.40%	Aerobic	BDH, 10139	(Gutnick et al., 1969)

2.3.9 Antimicrobial susceptibility testing of *Salmonella* isolated from venomous and non-venomous reptiles

Antimicrobial susceptibility was determined using a modified version of the European Committee on Antimicrobial Susceptibility Testing (EUCAST) disk diffusion method (Matuschek et al., 2014) using Mueller Hinton (Lab M, LAB039) agar plates and a DISKMASTER ™ dispenser (Mast Group, UK, MDD64). Inhibition zone diameters were measured and compared to EUCAST zone diameter breakpoints for Enterobacteriaceae (Andrews and Howe, 2011). Isolates were first tested with six commonly used antibiotics (ampicillin 10 µg, chloramphenicol 30 µg, nalidixic acid 30 µg, tetracycline 30 µg, ceftriaxone 30 µg, and trimethoprim/sulfamethoxazole 25 µg) and were then tested with five additional antibiotics (meropenem 10 µg, gentamicin 10 µg, amoxicillin/clavulanic Acid 30 µg, azithromycin 15 µg, and ciprofloxacin 5 µg) if any resistance was seen to the primary antibiotics (all disks from Mast Group). If resistance was observed phenotypically, then the presence of antimicrobial resistance genes were investigated using the ResFinder software tool (Zankari et al., 2012). Antimicrobial resistance was defined as resistance to one antimicrobial to which isolates would normally be susceptible (Magiorakos et al., 2012).

2.4 Results and Discussion

2.4.1 Similarities in *Salmonella* prevalence between captive venomous snakes and non-venomous reptiles

Salmonella carriage is well documented amongst reptiles (**Table 2.1**), however, no published study reports the incidence of *Salmonella* in venomous snakes. The period prevalence of *Salmonella* was assessed in a collection of 106 venomous snakes housed at the LSTM venom unit between May 2015 and January 2017. Please note that the sampling of faecal isolates was largely completed prior to the start of this project and does not form part of the thesis work. All subsequent work forms part of this thesis. A remarkably high proportion (91%; 97/106) of the faecal samples contained *Salmonella* (**Table S2.1**), which should be seen in the context of the significant carriage rate of *Salmonella* by other non-venomous reptiles described in the literature (Whiley et al., 2017). Variable rates of *Salmonella* carriage have been observed in collections of reptiles (**Table 2.1**), and the large proportion of venomous snakes carrying *Salmonella* in this study sits at the higher end of the reported spectrum. The findings pose important public health considerations for individuals who work with venomous snakes housed in captivity, which may previously have been overlooked.

2.4.2 The diversity of *Salmonella* highlights the possibility of local transmission events and the role of long-term shedding

To assess diversity, 87 venomous snake-derived *Salmonella* isolates, 27 non-venomous reptile-derived *Salmonella* isolates and one venomous reptile-derived *Salmonella* isolate were whole genome sequenced. *In silico* serotyping revealed 58 different *Salmonella* serovars (**Figure 2.1**). A wide range of serovars was found in each of the venomous and non-venomous snake collections. Given that many of the serovars identified have a very broad host specificity, there was no suggestion that the presence of particular serovars was linked to the venom status of the reptile.

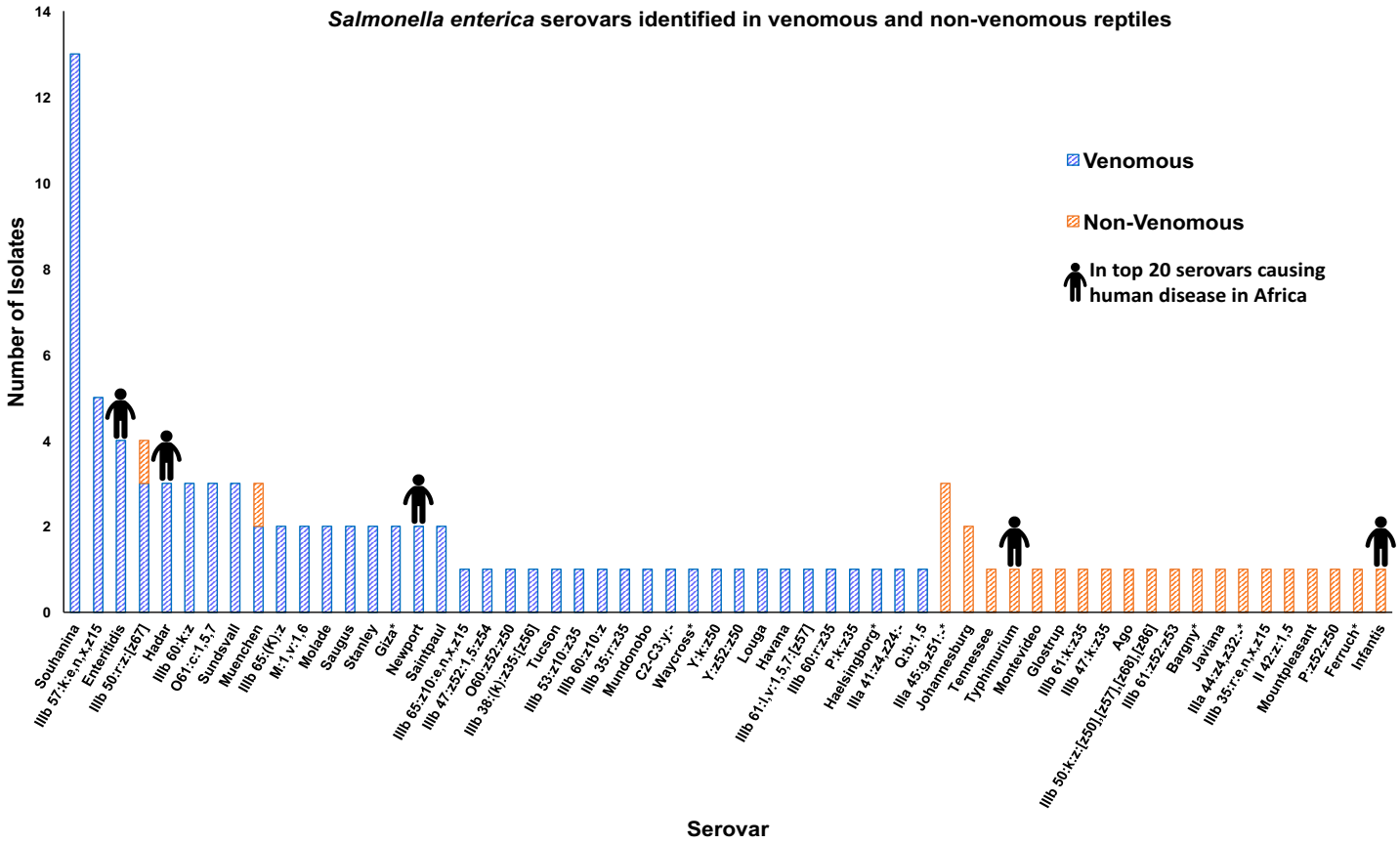


Figure 2.1 The distribution of the 58 *Salmonella enterica* serovars isolated from venomous and non-venomous reptiles.

Each bar represents the total number of isolates which belonged to each serovar. Serovars containing isolates that had multiple serovar designations from SISTR (**Section 2.3.4**) are indicated with asterisks. Human pictographs are displayed on serovars which are amongst the top 20 isolated from humans in Africa. Data is based on the global monitoring of *Salmonella* serovar distribution from the World Health Organisation (WHO) global foodborne infections network data (Hendriksen et al., 2011).

Similar levels of *Salmonella* carriage were seen in wild-caught and captive-bred reptiles. It is likely that the difference in serovar distribution reflects the sourcing of the reptiles from two independent housing facilities, and represents a limitation of this study. Nevertheless, a high level of *Salmonella* diversity was identified amongst both venomous and non-venomous reptiles. Following whole genome sequencing, multi-locus sequence typing was used for sub-serovar genetic characterisation of *Salmonella* (Alikhan et al., 2018; Maiden et al., 2013). In all cases, isolates falling within the same serovar had identical sequence types (**Table S2.1**), reflecting the intra-serovar homogeneity of the *Salmonella* isolated in this study.

The most common serovar to be identified amongst the venomous snake isolates was *S.* Souhanina (n = 12). All *S.* Souhanina isolates clustered locally on the phylogeny (**Figure 2.2**) falling within a 5 SNP cluster characteristic of a clonal expansion event (Dallman et al., 2018). Four of these isolates were found in captive-bred reptiles, whilst 11 isolates came from venomous snakes which originated in Cameroon, Uganda, Tanzania, Nigeria, Togo and Egypt. The close phylogenetic relationship between the *S.* Souhanina isolates that belong to the same MLST type (ST488) from imported animals with a range of origins and from captive animals suggests that local *Salmonella* transmission may be occurring. Local transmission of near-identical salmonellae could occur between snakes or as a result of a single contaminated food source such as frozen mice (Fuller et al., 2008; Kanagarajah et al., 2018).

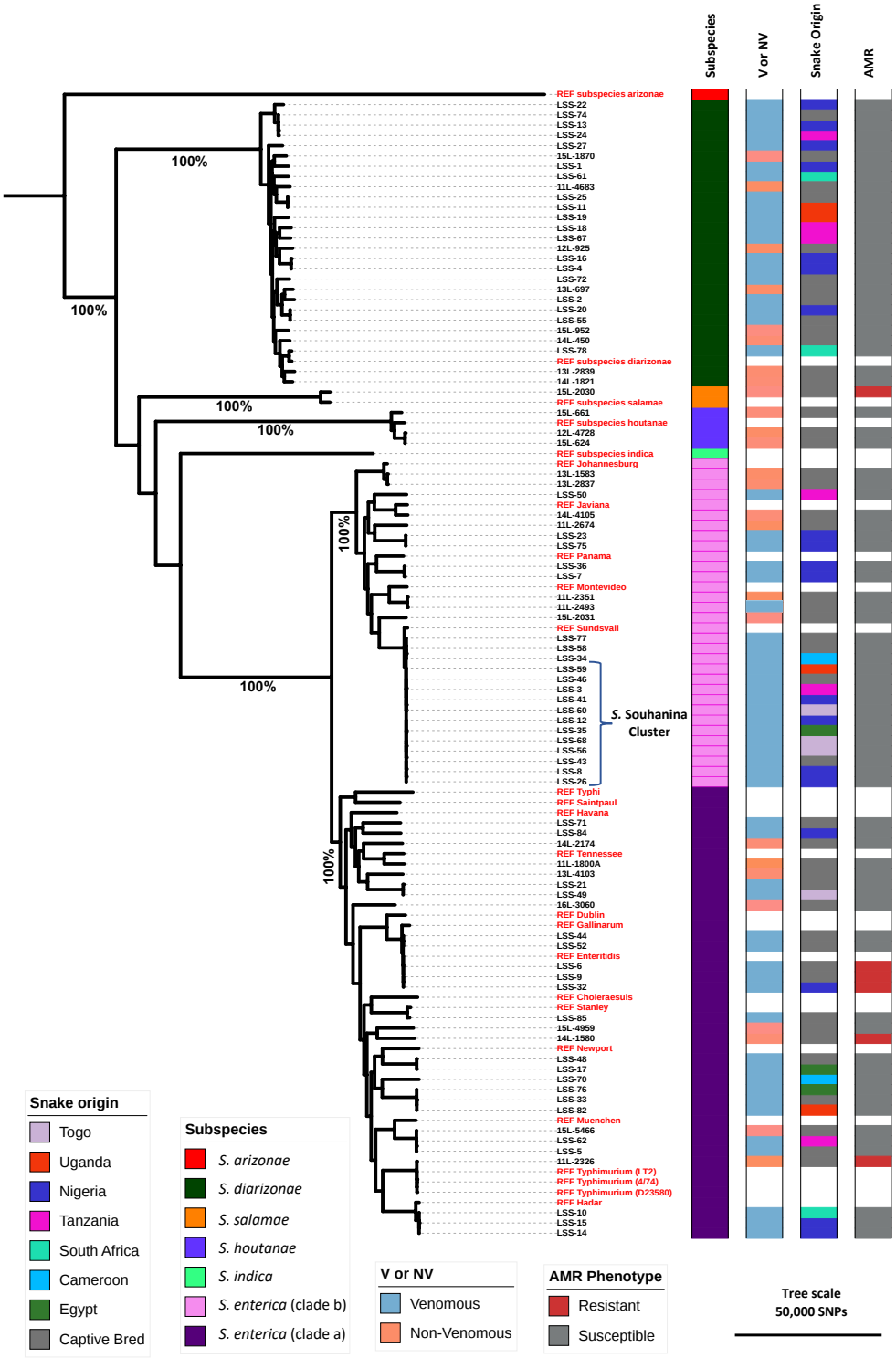


Figure 2.2 Diversity of *S. enterica* isolated from venomous snakes and non-venomous reptiles.

Core genome maximum likelihood phylogenetic tree (**Section 2.3.6**). The 25 contextual reference genomes representing previously sequenced isolates are indicated. Bootstrap values are indicated on key branches. Colour strips showing metadata are as follows; Subspecies – Subspecies of isolate, Origin – Country of origin of the snake, V or NV – Depicts whether the reptile host was venomous (V) or non-venomous (NV), AMR phenotype – Isolates shown in red were resistant to one or more antimicrobial agent (**Section 2.3.9**). M.L. Redway assisted with phenotypic AMR characterisation. Metadata for the contextual reference genomes appear as white. Tree was visualised using iTOL (*https://itol.embl.de)*.

Although the data suggested that *S.* Souhanina was locally transmitted within the herpetarium, a single source of *Salmonella* would not explain the wide variety of serovars and MLST types carried by the venomous snakes. Significant *Salmonella* diversity was reported in a study that involved 166 faecal samples from wild-caught reptiles in Spain, identifying 27 unique serovars (Briones et al., 2004). Another study looking at *Salmonella* diversity in wildlife in New South Wales, Australia identified 20 unique serovars amongst 60 wild-reptiles (Simpson et al., 2018). I speculate that the majority of the diversity of *Salmonella* identified originated from wild-caught reptiles and reflect their varied habitats in Africa.

Underpinning the strategy for sampling *Salmonella* was the assumption that venomous snakes can carry and shed *Salmonella* for long periods of time. The longitudinal shedding of *Salmonella* has been reported in 12 captive non-venomous snakes from 7 different species. Over 10 consecutive weeks, 58% of the snakes shed *Salmonella* intermittently (Goupil et al., 2012). Chronic *Salmonella* carriage has been reported in many other animals, including laying hens which shed the bacteria continually for up to 10 weeks (Gast et al., 2017). To assess the continuity of *Salmonella* shedding from venomous snakes in this study, faecal samples were collected from a Western Green Mamba from Togo over a three-month period between 31st October 2016 and 31st January 2017. All three faecal samples contained *Salmonella* which belonged to sequence type ST488, showing that individual snakes have the capacity to shed the same sequence type of *Salmonella* over a 90-day period in this study.

I conclude that the majority of the reptile-derived *Salmonella* described in this study were likely acquired by reptiles prior to captivity, whilst some isolates were transmitted locally within the herpetarium.

2.4.3 Venomous reptiles carry multidrug resistant *Salmonella* serovars of clinical relevance

Because the majority of venomous snakes examined in this study were of African origin or belonged to a species of snake native to the African continent (**Figure 2.2**), the *Salmonella* serovars isolated from all reptiles in this study were compared with those most frequently associated with human disease in Africa. The *Salmonella* serovar distribution was reported by the WHO global foodborne infections network data bank based on data from quality assured laboratories in Cameroon, Senegal and Tunisia (Hendriksen et al., 2011) (**Figure 2.1**). Eleven snake-derived isolates belonged to serovars commonly pathogenic in humans. This finding prompted me to determine the proportion of all venomous snakes and non-venomous reptiles that carried antimicrobial resistant *Salmonella* (**Table 2.4**). In *Salmonella* collected from venomous snakes, 4.1% of isolates (4/97) were resistant to at least one antimicrobial and two isolates were multidrug resistant (**Table 2.4**). Three resistant isolates from venomous snakes belonged to the serovar Enteritidis and were closely related to the global *S.* Enteritidis epidemic clade which causes human disease in Africa (Feasey et al., 2016). These findings demonstrate that venomous snakes are capable of carrying and shedding *Salmonella* that have the potential to cause disease in humans.

Table 2.4 Relating antimicrobial resistance to phenotype and genotype

Isolate	Serovar[#]	Reptile	Scientific Name	Venom Status	Origin	AMR Resistance Phenotype*	Resistance Genes and Mutations	Antibiotic Family
LSS-6	Enteritidis	Monocled cobra	*Naja kaouthia*	Venomous	Captive Bred	Ampicillin Nalidixic Acid	*blaTEM-1B* *gyrA* p.D87Y (GAC to TAC) *aph(3")-1b* *sul2*	Beta-lactam Quinolones Aminoglycoside Sulphonamide
LSS-9	Enteritidis	Monocled cobra	*Naja kaouthia*	Venomous	Captive Bred	Ampicillin Chloramphenicol Nalidixic Acid Tetracycline Cotrimoxazole Azithromycin	*blaTEM-1B* *catA2**** *gyrA* p.D87G (GAC to GGC) *tet(A)* *sul2* *aph(3")-1b* *aph(6)-1d***	Beta-lactam Phenicol Quinolones Tetracycline Sulphonamide Aminoglycoside
LSS-28	Stanley	Malaysian Spitting Cobra	*Naja siamensis*	Venomous	Captive Bred	Ampicillin Tetracycline	*blaOXA-1* *tet(A)**** *aph(4)-la* *aac(6')lb-cr* *aac(3_1Va)**** *aac(6')lb-cr* *sul2* *sul1* *catB3**** *ARR-3*	Beta-lactam Tetracycline Aminoglycoside Fluoroquinolone Sulphonamide Phenicol Rifampicin
LSS-32	Enteritidis	Black-Necked Spitting Cobra	*Naja nigricollis*	Venomous	Nigeria	Ampicillin Chloramphenicol Nalidixic Acid Tetracycline	*blaTEM-1B* *catA2**** *gyrA* p.D87G (GAC to GGC) *tet(A)* *aph(3")-1b* *aph(6)-1d**** *sul2*	Beta-lactam Phenicol Quinolones Tetracycline Aminoglycoside Sulphonamide
11L-2326	Typhimurium	Python	*Pythonidae*	Non-Venomous	Captive Bred	Ampicillin Chloramphenicol Nalidixic Acid Tetracycline Cotrimoxazole Azithromycin	*blaCARB-2* *floR* *gyrA* p.D87Y (GAC to TAC) *tet(G)* *sul1* *aadA2*	Beta-lactam Phenicol Quinolones Tetracycline Sulfonamide Aminoglycoside
14L-1580	Unknown	Carpet Python	*Morelia spilota*	Non-Venomous	Captive Bred	Nalidixic Acid Tetracycline	*gyrA* p.D87G (GAC to GGC) *Tet(A)*	Quinolones Tetracycline
15L-2030	II 42:z:1,5	Olive Tree Skink	*Dasia Olivacea*	Non-Venomous	Captive Bred	Tetracycline	No resistance genes identified with ResFinder	

[#] Serovar predicted from genome sequence with SISTR (Yoshida et al., 2016)

* Determined experimentally, as described in Materials and Methods (**Section 2.3.9**)

** Resistance genes and mutations were identified using ResFinder as described in Material and Methods (**Section 2.3.9**)

*** Less than 100% sequence identity on ResFinder

NB: Multidrug resistance is defined as resistance to greater than three antibiotics

2.4.4 Phylogenetic diversity and molecular epidemiology of *Salmonella* in venomous snakes and non-venomous reptiles demonstrated for the first time

Venomous snakes can shed *Salmonella.* The vast diversity of *Salmonella* has long been acknowledged in the literature (Alikhan et al., 2018). To study the diversity of reptile-associated *Salmonella* from an evolutionary perspective, 87 high quality whole genome sequences were obtained for a phylogenetic comparison that involved 24 contextual *Salmonella* genomes (**Section 2.3.6**). The 87 genomes represented 60 isolates from venomous snakes, 26 *Salmonella* isolates from non-venomous reptiles and 1 *Salmonella* isolate from a venomous reptile. Following a comprehensive comparative genomic analysis, a total of 405,231 core genome SNPs were identified that differentiated the 87 isolates, and were used to infer a maximum likelihood phylogeny (**Figure 2.2**). SNPs are a valuable marker of genetic diversity (Dallman et al., 2018), and the identification of hundreds of thousands of core-genome SNPs reflects a high level of genetic diversity among the reptile associated *Salmonella* isolates. The collection of reptile-derived *Salmonella* represented most of the known diversity of the *Salmonella* genus (Brenner et al., 2000), spanning four of the six *Salmonella enterica* subspecies: *diarizonae*, *enterica*, *houtanae* and *salamae.* Reptile-derived *S. enterica* subspecies *enterica* isolates were approximately equally distributed into two distinct phylogenetic clusters, known as clade A (58%) and clade B (48%) (den Bakker et al., 2011; Falush et al., 2006; Parsons et al., 2011) (**Figure 2.2**). No significant association was found between venom status and phylogenetic group (Odds Ratio = 1.1, Confidence Interval = 0.3-3.0, $\chi 2 = 0.02$, *p*-value = 0.4).

2.4.5 Genotypic and phenotypic conservation of infection-relevant carbon source utilisation and virulence-associated genomic regions sheds new light on *Salmonella* ecology

The unique collection of diverse *Salmonella* isolates was used to determine the phenotypic and genotypic conservation of infection-relevant properties and genomic elements. Whilst the reptile-associated *Salmonella* belonged to five evolutionary groups, the majority of isolates were classified as *S. diarizonae* or *S. enterica*. The clustering of *S. enterica* into two clades (A and B) has previously been inferred phylogenetically based on the alignment of 92 core loci (den Bakker et al., 2011; Falush et al., 2006). The biological significance of *S. enterica* clade A and clade B has been established as the two clades differ in host specificity, virulence-associated genes and metabolic properties such as carbon utilisation (Parsons et al., 2011). The genome sequences were used to expand upon pre-existing knowledge and determine

phenotypic and genotypic conservation of metabolic and virulence factors across *S. diarizonae* and *S*. *enterica* (clades A and B).

Although the majority of *Salmonella* serovars of public health significance belong to clade A, certain clade B serovars such as *S.* Panama have been associated with invasive disease (Chaudhuri et al., 2013; Kostiala et al., 1992). The clade B *S. enterica* generally carry a combination of two *Salmonella* genomic islands. The *Salmonella* Pathogenicity Island-18 encodes an intracellularly expressed pore forming hemolysin *hlyE* and the cytolethal distending toxin islet which includes the gene *cdtB* (Crump et al., 2004; den Bakker et al., 2011). An association with invasive disease has been suggested because the two islands have previously only been identified in *S. enterica* serovar Typhi and Paratyphi A, which cause bloodstream infections (Crump et al., 2004; den Bakker et al., 2011). The combination of *hlyE* and *cdtB* genes were present in all *S. diarizonae* and *S. enterica* clade B isolates in this study, but absent from all but one *S. enterica* clade A isolate (14L-2174). The significant proportion of reptiles which carried *S. enterica* clade B could partially explain the increased likelihood of reptile-associated salmonellosis involving invasive disease, compared to non-reptile-acquired salmonellosis (Murphy and Oshin, 2015).

To assess metabolic differences that distinguish *S. enterica* clade A, *S. enterica* clade B and *S. diarizonae*, 39 reptile isolates were phenotypically screened for the ability to catabolise a number of infection-relevant carbon sources (Feasey et al., 2016; Gutnick et al., 1969; Kröger and Fuchs, 2009; Nichols et al., 2011) (**Section 2.3.8**, **Table 2.2** and **Table 2.3**). The sampling encompassed isolates from a range of snakes, origins and serovars (**Section 2.3.8**). A summary of the results for phenotypic carbon utilisation and the presence of genes associated with the cognate metabolic pathway is shown in **Figure 2.3**.

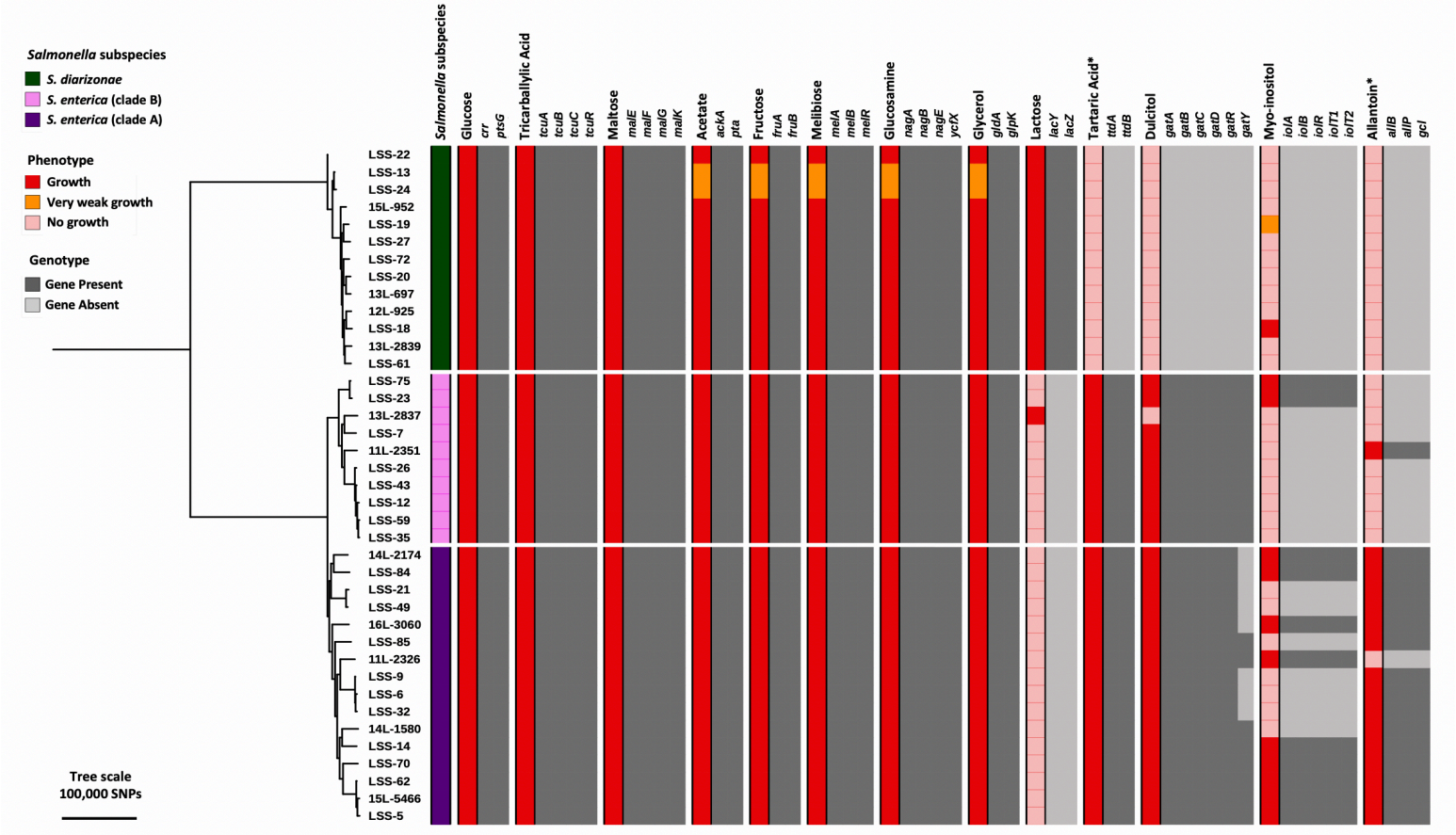


Figure 2.3 Phylogenetic context of carbon-source utilisation by reptile-derived-*Salmonella* isolates.

Carbon source utilisation data were mapped against the core genome phylogenetic tree. The maximum likelihood tree includes all reptile-derived *Salmonella* isolates which were assessed for carbon utilisation and had high quality genome sequences (see **Table 2.2**, **Table 2.3** and **Data S1**). Reference sequences for the majority of carbon utilisation and acquisition genes were taken from S. Typhimurium strain 4/74, and the *lac* gene sequences were from *E. coli* MG1655. Carbon sources which required anaerobic conditions are indicated with asterisk. E. V. Rodwell and H. J. Webster assisted in phenotypic characterisation of carbon sources. See **Section 2.3.8** for experimental details.

In general, the genotype accurately reflected phenotype in terms of carbon source utilisation; however, this was not always the case (**Figure 2.3**). Discrepancies between phenotypic growth and genotype suggests that mechanisms of *Salmonella* metabolism remain to be elucidated. For example, *S. diarizonae* isolate LSS-18 grew well on *myo*-inositol as a sole carbon source (**Figure 2.3**) but showed zero percent homology with any of the *iol* genes from the well-characterised *Salmonella* strain 4/74.

The ability to utilise lactose was a property of most *S. diarizonae* isolates, consistent with previous reports that 85% of *S. diarizonae* are *Lac*$^{+}$ (Lamas et al., 2018). It is estimated that less than 1% of all *Salmonella* ferment lactose due to the absence of the *lac* operon from the *S. enterica* subspecies (McDonough et al., 2000). It was interesting to discover that one non-venomous snake isolate (13L-2837) which belongs to *S. enterica* clade B was capable of utilising lactose as a sole carbon source. Isolate 13L-2337 belongs to the serovar *S.* Johannesburg and this is the first published occurrence of a Lac^{+} *S.* Johannesburg isolate. The 13L-2837 pan-genome had zero percent homology to the *lac* genes from reference strain *E. coli* MG1655 (sequence in S1 Text) (results in **Figure 2.3**), suggesting an alternative method for lactose utilisation. The 13L-2837 *S.* Johannesburg isolate also lacked the ability to grow on dulcitol, despite possessing all of the relevant *gat* genes, raising the possibility of an inverse relationship between the ability of *Salmonella* to utilise dulcitol and lactose as a sole carbon source. These findings require further investigation which is beyond the scope of the current study.

The majority of *S. enterica* clade A and clade B isolates utilised dulcitol, whereas dulcitol was rarely used as a sole carbon source by *S. diarizonae*. These findings are consistent with a study of *Salmonella* derived from Australian sleepy lizards, which demonstrated that dulcitol utilisation was observed in almost all *S. enterica* and *S. salamae* isolates but only 10% of *S. diarizonae* isolates (Parsons et al., 2015). Over 50% of the *S. enterica* clade A isolates lacked the *gatY* gene but grew well on dulcitol as a sole carbon source, suggesting that GatY is not required for dulcitol catabolism. A variety of repertoires of dulcitol catabolic genes have been described across *Salmonella*, with individual serovars carrying one of two *gat* gene clusters (Nolle et al., 2017). Both of these clusters carry the *gatY* gene. The findings may indicate that a third *gat* gene cluster is carried by some *Salmonella* serovars.

In the majority of cases, allantoin was only utilised as a sole carbon source by *S. enterica* clade A isolates, consistent with a previous report that described an association of clade A with the allantoin catabolism island (den Bakker et al., 2011). The majority of clade B isolates lacked the allantoin catabolism island and thus were unable to utilise allantoin as a sole carbon source. However, one clade B isolate was identified as an exception, isolate 11L-2351 which was sampled from a non-venomous reptile. This isolate belongs to the serovar Montevideo, which is frequently associated with outbreaks of human salmonellosis (Gaffga et al., 2012; Gieraltowski et al., 2013; Stocker et al., 2011). In reptiles, the end product of the purine catabolic pathway is not allantoin, but uric acid (den Bakker et al., 2011). The consequent absence of allantoin from the snake gastrointestinal tract could explain why a substantial number of *S. enterica* clade B were found in snakes.

It is possible that the gain and loss of allantoin catabolic genes is relevant to host-specificity. Allantoin is a metabolic intermediate derived from uric acid that is found at high levels in some animal hosts (chickens and rats) but not in humans (Matiasovicova et al., 2011). A relationship between the pseudogenisation of the allantoin metabolic genes and niche adaptation has also been proposed for the iNTS reference isolate for *S.* Typhimurium: D23580 (Kingsley et al., 2009; Okoro et al., 2015). Compared with *S.* Typhimurium isolate 4/74, which shows a broad host range, D23580 is unable to utilise allantoin as a sole carbon source, consistent with the adaptation of invasive *Salmonella* in Africa towards non-allantoin producing hosts (Kingsley et al., 2009; Okoro et al., 2015). Furthermore, accumulation of pseudogenes in the allantoin degradation pathway has been reported in host-restricted *Salmonella* serovars which cause invasive disease, suggesting that the ability to grow on allantoin is a marker of a switch from enteric to invasive disease (Langridge et al., 2015). These findings may reflect the clinical observation that snake-acquired salmonellosis is frequently an invasive disease that commonly results in hospitalisation, compared to disease caused by *Salmonella* derived from allantoin-producing hosts such as chickens.

2.6 Perspective

Although reptiles are known to harbour a diverse range of *Salmonella* bacteria, until now *Salmonella* carriage has not been examined in many key reptilian species. Here, I have shown that venomous snakes harbour and shed a wide variety of *Salmonella* serovars that represent much of the spectrum of the *Salmonella* genus and are phylogenetically distributed in a similar way to *Salmonella* found in non-venomous reptiles. I demonstrated that venomous snakes can carry and excrete *Salmonella* serovars which cause human disease. My findings raise the possibility that venomous snakes represent a previously uncharacterised reservoir for *Salmonella* both in captive settings and in the wider environment. Further study is required to investigate the relationship between clinical cases and reptile-derived *Salmonella* in tropical regions inhabited by venomous reptiles such as Africa. This study provides a good baseline for this future work.

Reptiles are an ideal population of animals for the study of genus-level evolution of *Salmonella* because they carry phylogenetically-diverse isolates that belong to the majority of *Salmonella* subspecies. By demonstrating the phenotypic and genotypic conservation of metabolic properties across three phylogenetic groups of *Salmonella* I have shed new light on the evolution of *Salmonella* serotypes.

All the reptile-derived genome-sequenced *Salmonella* isolates described in this chapter, are now in long-term cryogenic storage in the Hinton laboratory at the University of Liverpool, and are available for future research.

Chapter 3

The stepwise evolution of *Salmonella* Typhimurium ST313 responsible for bloodstream infection in Africa

3.1 Acknowledgement of the specific contribution of collaborators

Much of the content of this chapter is under review at Nature Microbiology. I acknowledge the following contribution of collaborators to the phenotypic experiments described in this chapter. Unless specified below, all work was completed by the author.

Blanca Perez-Sepulveda University of Liverpool, UK	Phenotypic characterisation of melibiose and tartaric acid usage by ST313 lineages. Phenotypic characterisation of RDAR morphology in ST313 lineages (**Section 3.4.7**).
Rocío Canals University of Liverpool, UK	Construction of the *S.* Typhimurium 4/74 Δ*lpxO*::*aph* mutant by λ Red recombineering (**Section 3.4.7**).
Jessica Bevington University of Liverpool, UK	Phenotypic characterisation of AMR profiles (**Section 3.3.10**).
Rebecca J. Bengtsson University of Liverpool, UK	Provided bioinformatic support and helped with troubleshooting and narrative direction.
Nicolas Wenner University of Liverpool, UK	Phenotypic characterisation of the pAnkS encoded LsoA/ LsoB toxin-antitoxin proteins which act as a bacteriophage exclusion system in *Salmonella* (**Section 3.4.3**).
Ella Rodwell University of Liverpool, UK	Phenotypic confirmation of plasmid and prophage repertoire (**Section 3.3.14**).
Benjamin Kumwenda University of Malawi, Malawi	Provided access to contemporary *S.* Typhimurium isolates from Malawi.
Xiaojun Zhu University of Liverpool, UK	Phenotypic characterisation of catalase activity in ST313 lineages (**Section 3.4.7**).
Rebecca J. Bennett University of Liverpool, UK	Provided bioinformatic support and helped with troubleshooting BEAST analysis and narrative direction.
George E. Stenhouse University of Liverpool, UK	Provided bioinformatic support and helped with troubleshooting BEAST analysis and narrative direction.
P. Malaka De Silva University of Liverpool, UK	Helped with narrative direction and reviewing results.
Hermione Webster University of Liverpool, UK	Phenotypic characterisation of tartaric acid usage by ST313 lineages (**Section 3.4.7**).
Jose A. Bengoechea Queens University Belfast, UK	Lipid A analysis by mass spectrometry (**Section 3.4.7**).
Amy Dumigan Queens University Belfast, UK	Lipid A analysis by mass spectrometry (**Section 3.4.7**).
Alicia Tran-Dien Institut Pasteur, France	Assisted with the collection of isolates from the Institut Pasteur.
Reenesh Prakash MLW, Malawi	Assisted with the collection of isolates from MLW.
Happy C. Banda MLW, Malawi	Assisted with the collection of isolates from MLW.
Lovemore Alufandika MLW, Malawi	Assisted with the collection of isolates from MLW.
Mike P. Mautanga MLW, Malawi	Assisted with the collection of isolates from MLW.
Arthur Bowers-Barnard University of Liverpool, UK	Assisted with the collection of isolates from MLW.
Alexandra Y. Beliavskaia University of Liverpool, UK	Long read sequencing of BKQZM9 (**Section 3.3.13**).
Alexander V. Predeus University of Liverpool, UK	Long read sequencing of BKQZM9 (**Section 3.3.13**).
Will Rowe University of Liverpool, UK	Provided bioinformatic support and helped with troubleshooting.

Alistair C. Darby University of Liverpool, UK	Long read sequencing of BKQZM9 (**Section 3.3.13**).
Neil Hall Earlham Institute, UK	Sequencing of isolates as part of the 10,000 *Salmonella* genomes project.
François-Xavier Weill Institut Pasteur, France	Conceptualisation, methodological design and provided access to samples at the Institut Pasteur.
Melita A. Gordon MLW, Malawi	Conceptualisation and provided access to samples at MLW.
Nicholas A. Feasey MLW, Malawi	Conceptualisation and methodology design.
Kate S. Baker University of Liverpool, UK	Supervision, conceptualization and methodology design.
Jay C.D. Hinton University of Liverpool, UK	Supervision, conceptualization and methodology design.

3.2 Introduction

Bloodstream infections caused by NTS have emerged as a major public health concern, resulting in an estimated 49,600 deaths every year in Africa (Stanaway et al., 2019a). iNTS disease disproportionally affects immunocompromised individuals such as adults with HIV and children under five years of age with malaria, malnutrition or severe anaemia living in lower-to-middle income countries (Feasey et al., 2012). Recently, a systematic review established that *Salmonella* was the most frequently isolated pathogen in hospitalised patients diagnosed with community-onset bloodstream infections in Africa and Asia (2008 to 2018) (Marchello et al., 2019). *S.* Typhimurium is responsible for approximately two thirds of iNTS disease cases in Africa (Reddy et al., 2010), and consequently is a focal point for research.

The majority of *S.* Typhimurium isolated from bloodstream infections in Africa belong to sequence type ST313, a single locus variant of *S.* Typhimurium ST19 (Kingsley et al., 2009; Okoro et al., 2015, 2012). ST19 is a common cause of gastrointestinal infection globally and can cause bloodstream infections in immunocompromised individuals (Majowicz et al., 2010). Phylogenetic reconstruction has revealed that two distinct and tightly clustered lineages of ST313 exist in Africa, differing from each other at a whole genome sequence level by 455 single nucleotide polymorphisms (SNPs) (Kingsley et al., 2009). ST313 L1 and ST313 L2 have been estimated to have emerged ~52 and ~35 years ago respectively, and differ from ST19 by over 700 SNPs (Okoro et al., 2012). These African ST313 L1 and L2 are genetically distinct from ST313 found in the United Kingdom and Latin America (Almeida et al., 2017; Ashton et al., 2017).

It is proposed that antimicrobial resistance has contributed to the success of ST313 lineages in Africa, with a large proportion of strains being multidrug resistant, specifically against ampicillin, chloramphenicol and trimethoprim/sulfamethoxazole (henceforth cotrimoxazole) (Feasey et al., 2015). Amongst *S*. Typhimurium ST313, chloramphenicol resistance is specifically found in ST313 L2, and is thought to have contributed to the clonal replacement of ST313 L1 by ST313 L2 prior to 2001, when chloramphenicol was the empirical treatment choice for suspected sepsis in Malawi (Okoro et al., 2012). Subsequently, chloramphenicol was replaced by the oral fluoroquinolone, ciprofloxacin, for treatment of iNTS infections from approximately 2002 onwards in several parts of Africa including Malawi

(Feasey et al., 2015). The third-generation cephalosporin, ceftriaxone became the empirical therapeutic for suspected sepsis from 2005 (Feasey et al., 2015). In 2014, an XDR isolate of ST313 L2 was identified in Malawi with ESBL-mediated resistance to third-generation cephalosporins and resistance to fluoroquinolones including ciprofloxacin (Feasey et al., 2014). More recently, an XDR sub-lineage of ST313 L2 (ST313 L2.1) was identified in the DRC with additional resistance to azithromycin (Van Puyvelde et al., 2019). ST313 L2.1 also displays chromosomal evidence of further evolution within ST313, highlighting the importance of continued surveillance of *S.* Typhimurium strains causing bloodstream infection. The MDR phenotypes limit the available therapeutic options, prompting the need for an effective vaccine to prevent iNTS disease (Tack et al., 2020).

In addition to AMR, loss-of-function mutations (genome degradation) have also featured in the ST313 lineages (Piccini and Montomoli, 2020). Evidence of host adaptation in *S.* Typhimurium ST313 involved specific pseudogenes that inactivated gene function and modified metabolic, structural and virulence phenotypes (Bogomolnaya et al., 2013; Canals et al., 2019; Carden et al., 2017; Kingsley et al., 2009; Okoro et al., 2015; Singletary et al., 2016). A similar pattern of genome degradation is found in other invasive bacteria with restricted host ranges such as *Yersinia pestis* (Chain et al., 2004)*, Mycobacterium leprae* (Cole et al., 2001) and *S.* Typhi (Holt et al., 2009). Notably however, *S.* Typhimurium ST313 retains the ability to cause diarrhoea in humans (Lokken et al., 2016), and so has not undergone the same level of niche-specialisation as *S.* Typhi.

The functional relevance of genome degradation in ST313 has been investigated at the transcriptomic, proteomic and phenotypic levels for several key isolates (Aulicino et al., 2018; Bogomolnaya et al., 2013; Canals et al., 2019; Carden et al., 2017; Kingsley et al., 2009; Okoro et al., 2015; Singletary et al., 2016). To understand the evolutionary trajectory and epidemiological relevance of functional degradative events, a population-level study of *Salmonella* genomes from a wide range of contemporary and historical isolates is required. Here, I studied a large and up-to date collection of ST313 to analyse the stepwise evolution of *S.* Typhimurium causing bloodstream infection in Africa. I uncovered a pan-susceptible ST313 lineage (ST313 lineage 3 (L3)) which emerged in Malawi in 2016, with a distinct genome degradation pattern. Despite only appearing recently in clinical surveillance, my core and accessory genome analyses revealed ST313 L3 to be a phylogenetic intermediate

between ST313 L1 and ST313 L2. This facilitated my genotypic characterisation of the crucial functional degradative events that occurred along the evolutionary pathway of invasive *S.* Typhimurium across Africa.

3.3 Methods

3.3.1 Dataset

Salmonella samples were derived from two archived blood culture collections. The main dataset was sourced from the Malawi-Liverpool Wellcome Clinical Research Programme (MLW) in Blantyre, Malawi which contains over 14,000 *Salmonella* bloodstream isolates from patients with iNTS disease at the Queen Elizabeth Central Hospital between 1996 and 2018. Isolates listed as *S.* Typhimurium were extracted from the metadata file for the entire collection and stratified by AMR profile into the following four categories: susceptible, resistant to one first line agent, resistant to two first line agents or MDR. First line agents were considered to be ampicillin, cotrimoxazole and chloramphenicol. Random sub-sampling was then performed using random number generator (RAND function in Excel, Microsoft) and selected samples were collected and resuscitated from the freezer archives at MLW. It was possible to recover 647 isolates for whole-genome sequencing.

To complement the MLW data, a contextual dataset was sourced from the Institut Pasteur (IP), Paris. The dataset consisted of 96 *S.* Typhimurium isolates collected from human extraintestinal sites from patients contaminated in Algeria, Burkina Faso, Cameroon, Central African Republic, Congo, France, Côte d'Ivoire, Madagascar, Mali, Niger, Senegal, Sudan, Togo and Vietnam between 1966 and 2012. All 96 *S.* Typhimurium samples were selected, resuscitated and sent for whole genome sequencing.

3.3.2 Whole genome sequencing of short reads

Isolates were prepared by inoculating a single bead of frozen *Salmonella* stock into a FluidX 2D Sequencing Tube (FluidX Ltd, UK) containing 100 µL of buffered peptone water (Oxoid, CM0509) and incubating overnight at 37°C. Cultures were heat killed in a 95°C water bath for 20 minutes to produce thermolysates, prior to DNA extraction (MagAttract kit, Qiagen) and whole-genome sequencing, as part of the 10,000 *Salmonella* Genomes Project (https://10k-salmonella-genomes.com/) (Perez-Sepulveda et al., 2020). Illumina Nextera XT DNA Libraries were prepared (Illumina, FC-131-1096) and sequenced (Illumina HiSeq 4000) in multiplex (768) as 150 bp paired-end reads.

3.3.3 Assembly and annotation of short reads

Trimmomatic (Bolger et al., 2014) v0.36 was used to trim adapters and Seqtk v1.2-r94 (https://github.com/lh3/seqtk) was used to trim low-quality regions using the trimfq flag. FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC v1.0 (http://multiqc.info) were used for quality control of sequence reads according to the following criteria: passed basic quality statistics, per base sequence quality, per base N content, adapter content and an average GC content of between 47 % and 57 %. Only high-quality reads were used in downstream analysis. Unicycler (Wick et al., 2017) v0.3.0 was used to assemble genomes and QUAST (Gurevich et al., 2013) v4.6.3 was used to evaluate assembly quality to standards consistent with EnteroBase (Alikhan et al., 2018). Specifically, N50 > 20 kb, 600 or fewer contiguous sequences, total number of bases between 4Million base pairs (Mbp) and 5.8Mbp. Prokka (Seemann, 2014) v1.12 was used to annotate the genomes.

3.3.4 Sequence typing

Serotyping was confirmed using SISTR (Yoshida et al., 2016) v1.0.2 and compared with the original metadata table (**Table S3.1**). The strains were assigned a Multi Locus Sequence Type using MLST (Larsen et al., 2012) v2.10 based on the conservation of seven housekeeping genes.

3.3.5 Reference mapping to D23580

Trimmed sequencing data were mapped against the *S.* Typhimurium ST313 L2 reference genome D23580, using BWA mem (Li and Durbin, 2009) v0.7.10-r789. The reference genome consists of the D23550 chromosome (GenBank accession: LS997973.1), and four plasmids including pSLT-BT (GenBank accession: LS997974.1), pBT1 (GenBank accession: LS997975.1), pBT2 (GenBank accession: LS997976.1) and pBT3 (GenBank accession: LS997977.1). Mapped reads were then cleaned and sorted using the SAMtools suite (H. Li et al., 2009) v1.7. Reads were realigned against the reference using GATK (McKenna et al., 2010) v3.7 by creating targets for realignment (RealignerTargetCreator) and performing realignment (IndelRealigner). Removal of optical duplicates was completed using Picard v2.10.1-SNAPSHOT (https://broadinstitute.github.io/picard/). Sequence variants were called using Bcftools v1.9-80 (http://samtools.github.io/bcftools) to generate a reference-based pseudogenome for each sample with greater than 10x depth. QualiMap (García-

Alcalde et al., 2012) v2.0 was used to identify a mean sample depth of 35.17x across all isolates.

3.3.6 Phylogenetic reconstruction of dataset

High quality pseudogenomes (MLW *n* = 608, contextual *n* = 72) were concatenated, plasmids were excluded and Gubbins (Croucher et al., 2014) v2.2 was used to remove recombinant regions and invariable sites. The resultant multiple sequence alignment of reference-based pseudogenomes (12,013 variant sites) was used to infer a maximum likelihood phylogeny using RAxML-ng (Stamatakis, 2014) v0.6.0 with 100 bootstrap replicates to assess support. To assign clusters, rhierBAPs (Tonkin-Hill et al., 2018) was used specifying two cluster levels, 20 initial clusters and infinite extra rounds. Visualisations were performed using iTOL (Letunic and Bork, 2016) v4.2.

3.3.7 Phylogenetic reconstruction of dataset and publicly available dataset

For contextual interpretation, sequencing data from 286 previously published *S.* Typhimurium ST313 were retrieved from public repositories (**Table S3.1**) and constructed in a core gene phylogeny. Roary (Page et al., 2015) v3.11.0 was used to generate a core gene alignment and SNP-sites (Page et al., 2016) v2.3.3 was used to extract SNPs. The resultant multiple sequence alignment (15,240 variant sites) was used to construct a maximum likelihood phylogeny using RAxML-ng (Stamatakis, 2014) v0.6.0 with 100 bootstrap replicates to assess branch support. Visualisations were made using iTOL (Letunic and Bork, 2016) v4.2.

3.3.8 Temporal phylogenetic reconstruction

To determine the evolutionary history of ST313, a chronogram was produced using Bayesian phylogenetic inference. Because the MLW collection is substantially larger than the contextual collection and to reduce computation time, 150/680 samples were chosen for inclusion in the analysis. Sampling reflected original sample selection for this study and included all contextual isolates which passed quality control (*n*=72) and a subset of MLW isolates randomly selected using a random number generator (*n*=78, Microsoft Excel). A reference-mapped multiple sequence alignment (7,231 variant sites) was created as described above. The alignment was loaded into BEAUti (Bayesian Evolutionary Analysis Utility)(Bouckaert et al., 2019) v2.6.1 and tip dates were extracted from the metadata file. The BEAST model test (Bouckaert and Drummond, 2017) was used to integrate over all

possible site models. The clock model was specified as “relaxed clock log normal” in line with previously published Bayesian analysis for *S.* Typhimurium (Ashton et al., 2017; Okoro et al., 2012; Van Puyvelde et al., 2019).

BEAST2 (Bouckaert et al., 2019) v2.6.1 was used to execute three independent chains of length 250,000,000, logging every 1,000 with 10% burn-in and accounting for invariant sites (Data S3.1). Tracer (Rambaut et al., 2018) v1.7.1 was used to assess convergence, with all parameter effective sampling sizes being > 200. LogCombiner (Bouckaert et al., 2019) v2.6.1 was used to combine tree files, resampled at a lower frequency of 10,000 and DensiTree (Bouckaert, 2010) v2.2.7 was used for visualisation. Finally, a maximum clade credibility tree was created using TreeAnnotator (Bouckaert et al., 2019) v2.6.0.

3.3.9 Determination of invasiveness index

The invasiveness index of each isolate was calculated using previously defined methods (Wheeler et al., 2018). Specifically, samples were analysed with the model described in Wheeler *et al.* (2018) using 196 top predictor genes for measuring invasiveness in *S. enterica.* The distribution of invasiveness index values for each lineage were compared using the Wilcoxon Mann Whitney test implemented through R (R Core Team, 2017) v3.4.0. A custom-made database of the top 196 invasiveness predictor genes was then created from the multi-fasta file provided in Wheeler *et al.* (2018). SRST2 (Inouye et al., 2014) v0.2.0 was used to flag any genes with mutations in ST313 L3 compared with ST313 L2. Details of specific mutations were investigated manually using MegaX (Kumar et al., 2018) v0.1.

3.3.10 Antimicrobial resistance testing and statistical analysis

Genetic determinants for AMR were identified using staramr v0.5.1 (https://github.com/phac-nml/staramr) against the ResFinder (Zankari et al., 2012) and PointFinder (Zankari et al., 2017) databases. Phenotypic antimicrobial susceptibility testing was performed using the EUCAST disk diffusion method (Matuschek et al., 2014) assisted by J. Bevington (University of Liverpool, UK). Antibiotic disks were ampicillin 10 µg, chloramphenicol 30 µg and trimethoprim/sulfamethoxazole 25 µg (Mast Group). To compare the results of genotypic against phenotypic testing, sensitivity and specificity were calculated for first line agents used over the study period using MedCalc’s test evaluation calculator with Clopper-Pearson confidence intervals (https://www.medcalc.org/calc/diagnostic_test.php).

3.3.11 Genomic conservation of plasmids, prophages and pseudogenes

A literature search was used to identify genomic regions known to vary between ST19 and ST313, including plasmids (Kingsley et al., 2009), prophages (Owen et al., 2017) and pseudogenes (Ashton et al., 2017; Bogomolnaya et al., 2013; Canals et al., 2019; Carden et al., 2017; Lawley et al., 2006; Nishino et al., 2006; Okoro et al., 2015; Singletary et al., 2016; Yang et al., 2015). A custom-made database of the identified regions was created, along with known variants. Sequences downloaded using the online tool SalComD23580 (http://bioinf.gen.tcd.ie/cgi-bin/salcom_v2.pl?_HL) (Canals et al., 2019). SRST2 (Inouye et al., 2014) v0.2.0 were then used to identify alleles present and the non-exact matches were investigated manually using MegaX (Kumar et al., 2018) v0.1. Results were manually mapped onto the phylogeny and chronogram using iTOL (Letunic and Bork, 2016) v4.2 which allowed inference of branches in which functional gene loss and plasmid and prophage acquisition occurred. Phenotypic testing was conducted to determine the relevance of plasmid and functional gene loss by B. Perez-Sepulveda, R. Canals, N. Wenner, X. Zhu, H. Webster (University of Liverpool, UK), J. Bengoechea and A. Dumigan (Queens University Belfast, UK).

3.3.12 Phylogenetic reconstruction of pSLT plasmid

Plasmid sequences were extracted from high quality pseudogenomes and snp-sites (Page et al., 2016) was used to extract SNPs. The resultant multiple sequence alignment of length 1,034 sites was used to infer a maximum likelihood phylogeny using RAxML-ng (Stamatakis, 2014) v0.6.0 with 100 bootstrap replicates to assess support. The resultant phylogeny was visualised in iTOL (Letunic and Bork, 2016) and convenience rooted to display lineages.

3.3.13 Long read sequencing of ST313 L3 reference strain BKQZM9

ST313 L3 strain BKQZM9 was selected for long read sequencing, based on having good overall quality statistics based on short read sequencing in terms of number of contigs and coverage. An Oxford Nanopore MinION (Jain et al., 2016) 9.4.1 flowcell and SQK-RAD004 rapid sequencing kit was used with base calling by Guppy v3.1.5 (https://nanoporetech.com/nanopore-sequencing-data-analysis#). Long read sequencing was performed by A. Beliavskaia, A. Predeus and A. Darby (University of Liverpool, UK). Approximately 120x coverage (600 Mb, 60,164 reads, read N50 33 kb) was generated. Hybrid genome assembly with the Illumina reads was done using Unicycler (Wick et al., 2017) v0.4.4

which generated a 4.9 Mb circular chromosome (GenBank Accession: CP060169), a 94 kb circular plasmid (pSLT) (Accession: CP060170) and a 1,975 bp circular plasmid (pBT3) (Accession: CP060171). Prokka (Seemann, 2014) v1.12 was used for annotation.

3.3.14 Genomic comparison of ST313 lineages

To determine the basic architecture of the ST313 L3 BKQZM9 genome, the identity of the chromosomal and plasmid regions were compared with existing sequences from ST313 L1 (A130), ST313 L2 (D23580) and ST313 UK (U2) using BLAST (Altschul et al., 1990) and visualised using the Artemis Comparison Tool (ACT) (Carver et al., 2005) v13.0.0. Putative prophage and plasmid sequences were extracted and identified using a BLAST (Altschul et al., 1990) search against known ST313 prophages and plasmids and the non-redundant NCBI nucleotide database. Prophage and plasmid presence/absence were also determined experimentally by PCR by E. Rodwell (University of Liverpool).

3.4 Results and Discussion

3.4.1 Assembling an informative collection of *S*. Typhimurium isolates

A combination of historical and contemporary *S.* Typhimurium bloodstream isolates were sampled from two sources (**Figure 3.1**). The main dataset consisted of 608 human bloodstream isolates from the Malawi-Liverpool Wellcome Trust Clinical Research Programme (hereafter referred to as MLW isolates) and their associated epidemiological metadata (**Table S3.1**). These were sampled from the MLW collection which contains approximately 8,000 *S.* Typhimurium isolates from patients presenting to the Queen Elizabeth Central Hospital in Blantyre (Malawi) between 1996 and 2018, and represents one of the most comprehensive blood culture archives available in sub-Saharan Africa. To improve sampling precision and ensure a range of AMR phenotypes were collected, the entire MLW dataset was stratified by AMR phenotype (**Section 3.3.1**) and 1,000 strains were randomly selected for inclusion in the study. The majority (647) of these isolates were resuscitated and whole-genome sequenced as part of the 10,000 *Salmonella* genomes project (Perez-Sepulveda et al., 2020). After assessing sequence quality (**Table S3.2**), 608 samples were included in the final dataset.

For context, 72 human bloodstream isolates collected between 1966 and 2012 by the Unité des Bactéries Pathogènes Entériques, Institut Pasteur, Paris, France (hereafter called the contextual collection) were also included. These were derived from 13 African, 1 European and 1 Asian country. The study periods were selected based on availability of isolates, and encompassed an epidemic of bloodstream infections caused by *S.* Typhimurium ST313 in Africa between 2002 and 2008 (Feasey et al., 2015). In total, 96 contextual isolates were whole-genome sequenced, 72 of which passed quality control (**Table S3.2**) and were included in the final dataset.

A description of the isolates is summarised in **Figure 3.1** and is available with all metadata and genome accession numbers in **Table S3.1**.

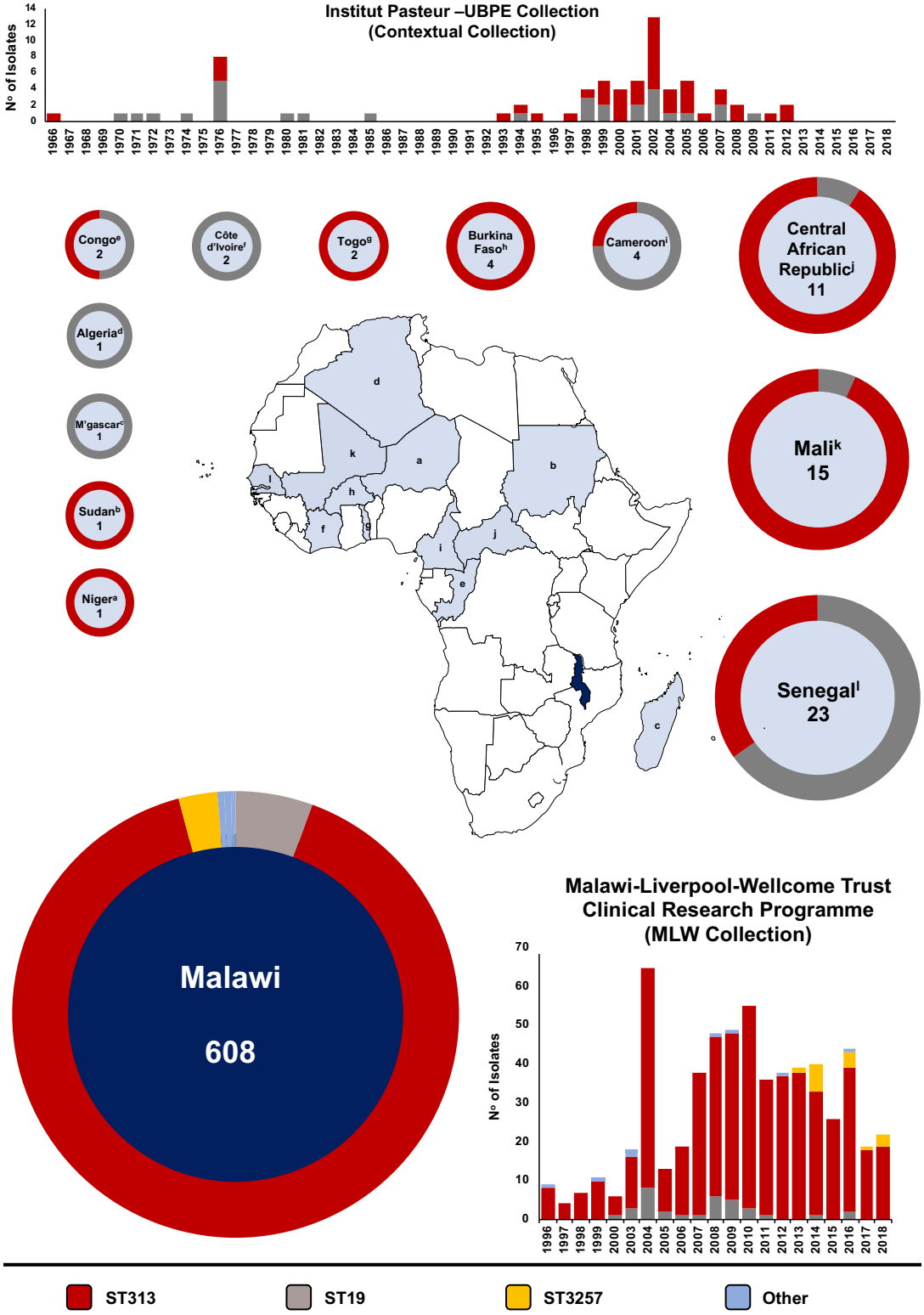


Figure 3.1 Bloodstream isolates of *S.* Typhimurium used in this study

Isolates were collected from either the Unité des Bactéries pathogènes entériques (UBPE) of the Institut Pasteur Centres (light blue) or the Malawi-Liverpool Wellcome Trust Clinical Research Center (dark blue) (**Section 3.3.1**). Bar graphs show the number of isolates of different sequence types collected by centre per year. Donut charts indicate the proportion of sequence types collected per country and show the total number of isolates from each location in the centre. Note that Madagascar is written as M'gascar. Letters appearing in superscript relate to location on map. The colour code is shown at the bottom of the figure, with red indicating *S.* Typhimurium ST313.

3.4.2 Population structure of *S.* Typhimurium

To define the population structure of *S.* Typhimurium that is currently causing bloodstream infection in Africa, the relationship of the 680 *S.* Typhimurium genomes from the MLW and contextual collections were explored. *In silico* prediction methods revealed that the two predominant sequence types causing bloodstream infection in Malawi were ST313 (*n*=549) and ST19 (*n*=34). Additionally, a small number of isolates (*n*=25) were typed as one of five single locus variants of ST313 (ST3257 [*n*=17], ST2080 [*n*=3], ST302 [*n*=3], ST4200 [*n*=1], ST4274 [*n*=1]). The contextual dataset contained two different sequence types: ST313 (*n*=43) and ST19 (*n*=29). Taken together, the ST313 isolates originated from ten African countries (**Figure 3.1**). It should be noted that there were less than five isolates that originated from six of these countries. Previous genome-based analyses of the ST313 epidemic in Africa have included isolates from Eastern and Central Africa including Malawi, Kenya, Uganda, DRC and Mozambique (Okoro et al., 2012; Van Puyvelde et al., 2019). Here, by including a range of countries, in particular from Western Africa, the known geographic range of ST313 has been expanded to Cameroon, Central African Republic, Niger, Senegal, Sudan and Togo (Almeida et al., 2017; Ashton et al., 2015; Jacob et al., 2019; Kagambega et al., 2018; Okoro et al., 2012; Tran-Dien et al., 2018; Van Puyvelde et al., 2019).

To investigate population structure, a whole-genome SNP-based maximum likelihood phylogenetic tree was constructed and cluster designation was performed using rHierBAPs (Tonkin-Hill et al., 2018). This analysis revealed three major ST313 clusters, four major ST19 clusters, and a small number of miscellaneous strains (**Figure 3.2**). Multi-locus sequence type variants formed discrete sub-lineages. The three ST313 clusters identified were highly clonal, with Cluster 1 and Cluster 2 corresponding to previously defined lineages (ST313 L1 and ST313 L2) (Kingsley et al., 2009). L1 was almost entirely clonally-replaced by L2 in the early 2000s, consistent with previous studies (Kingsley et al., 2009).

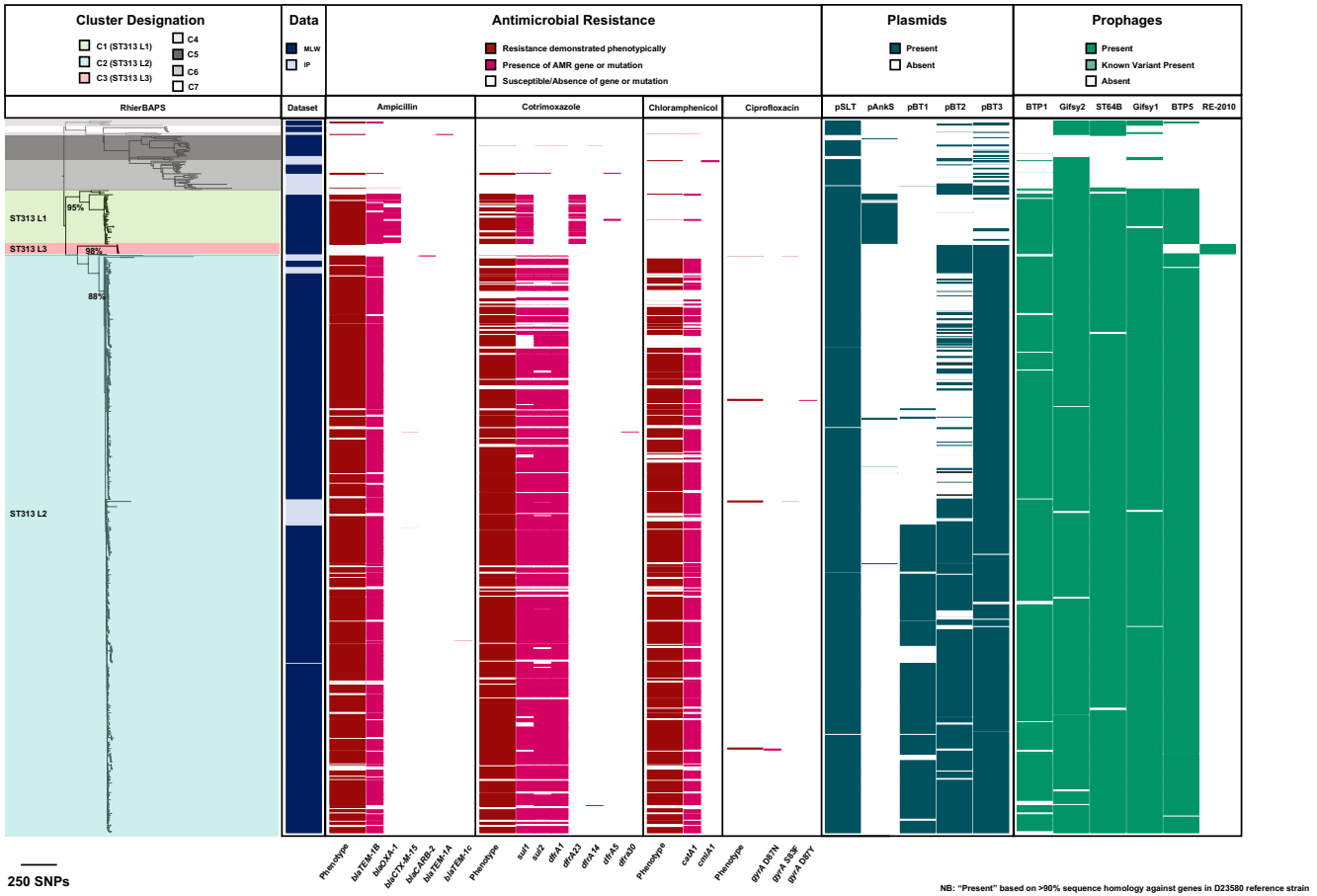


Figure 3.2 Phylogenetic relationship between all *S.* Typhimurium isolates in this study in the context of antimicrobial resistance and accessory genome

Maximum likelihood phylogeny based on the core genome SNP alignment of strains from Malawi-Liverpool Wellcome Clinical Research Center (MLW) and Institut Pasteur Unité des Bactéries pathogènes entériques Contextual Collection (IP) (**Section 3.3.6**). Background shading represents cluster designation (rHierBAPs), additional metadata is represented as adjacent colour strips. Bootstraps are displayed for branches leading to ST313 L1, L2 and L3. Visualised using iTOL (Letunic and Bork, 2016), rooted on ST19.

The genomes of contemporary isolates revealed that a third lineage has been circulating in Malawi since 2016 (hereafter called L3). To fully contextualise ST313 L3, a maximum likelihood phylogenetic tree was constructed that included all previously published ST313 genomes (**Table S3.3**) alongside the current dataset (**Figure 3.3**). This global contextualisation revealed that ST313 L3 is a novel lineage that forms a monophyletic cluster within a group of ST313 strains isolated in the United Kingdom and Brazil (Almeida et al., 2017; Ashton et al., 2017), raising the possibility of an international transmission event.

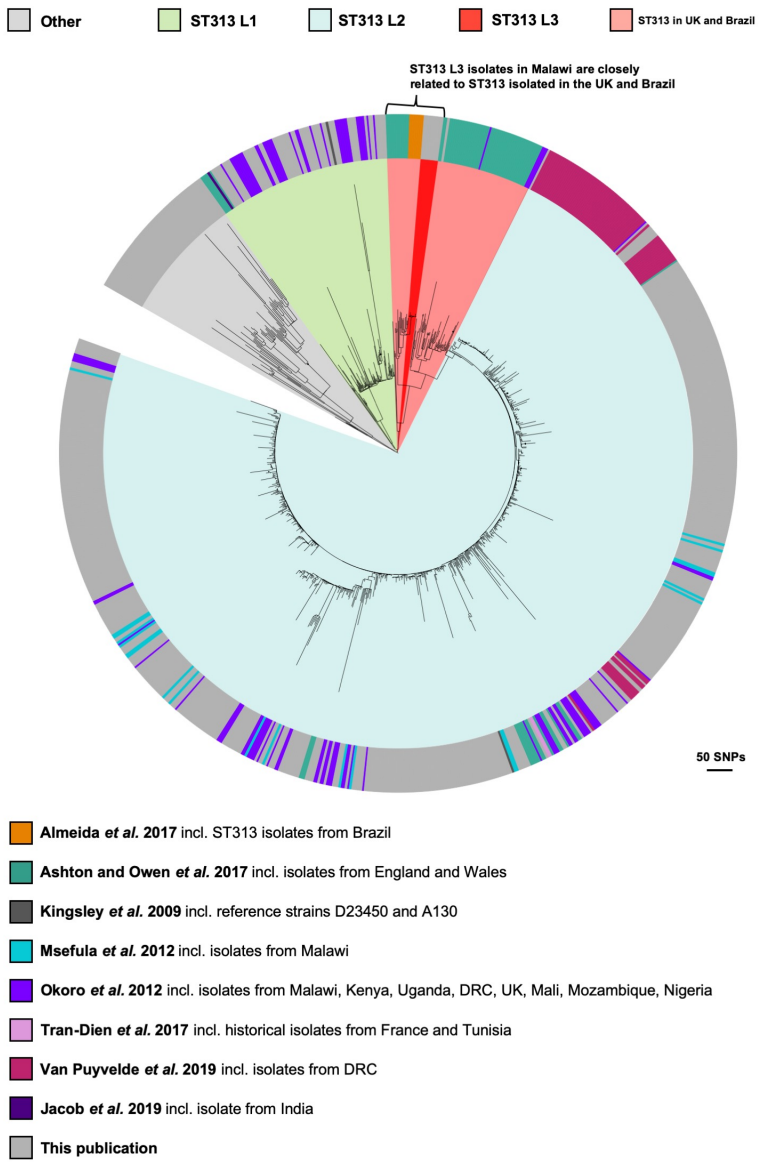


Figure 3.3 Phylogeny of *S*. Typhimurium ST313 isolates dating from 1966 to 2018

Maximum likelihood phylogeny based on the core gene SNP alignment of strains in this study in the context of published ST313 genomes (methods described in **Section 3.3.7**). Background shading on phylogeny represents cluster designation (rHierBAPs). Outer ring provides details of the relevant publication. Grey represents genomes reported in this thesis. Figure visualised using iTOL (Letunic and Bork, 2016) and rooted on ST19.

3.4.3 Accessory genome of *S.* Typhimurium ST313 lineages

Characterisation based on long-read sequencing showed that ST313 L3 strain BKQZM9 had a circular chromosome (4.8 Mb) with 99.98 % sequence identity to the ST313 L2 reference strain D23580 (Kingsley et al., 2009). The number of genes shared between D23580 and BKQZM9 was 4,529. An additional 167 genes were BKQZM9-specific, whereas 309 genes were exclusive to D23580, largely explained by differences in the prophage and plasmid repertoire of the two strains.

To understand the relationship between ST313 L3 and other ST313 lineages, the accessory genome was examined in the context of population structure (**Figure 3.4**). The plasmid profiles of the ST313 lineages were established. All lineages carried the virulence plasmid pSLT (Rotger and Casadesus, 1999), and ST313 L2 additionally contained plasmids pBT1, pBT2 and pBT3, consistent with previous studies (Kingsley et al., 2009). ST313 L1 was also found to carry a previously unreported 8,274 bp plasmid in the current study which shared 99.27 % sequence identity with pAnkS (Sahin et al., 2008) (GenBank NC_010896.1) and encoded the LsoA/ LsoB toxin-antitoxin proteins. This pAnkS-like plasmid was experimentally confirmed to carry a bacteriophage exclusion system by N. Wenner (University of Liverpool, UK) (Otsuka and Yonesaki, 2012). The ST313 L3 strain BKQZM9 contained the large 95 kb plasmid (pSLT) and the 1,975 bp pBT3 plasmid. A fourth plasmid corresponding to the 2,556 bp pBT2 was also identified using short read data. Thus, the plasmid profile of ST313 L3 shared more similarity to ST313 L2 than any other lineage (**Figure 3.4**).

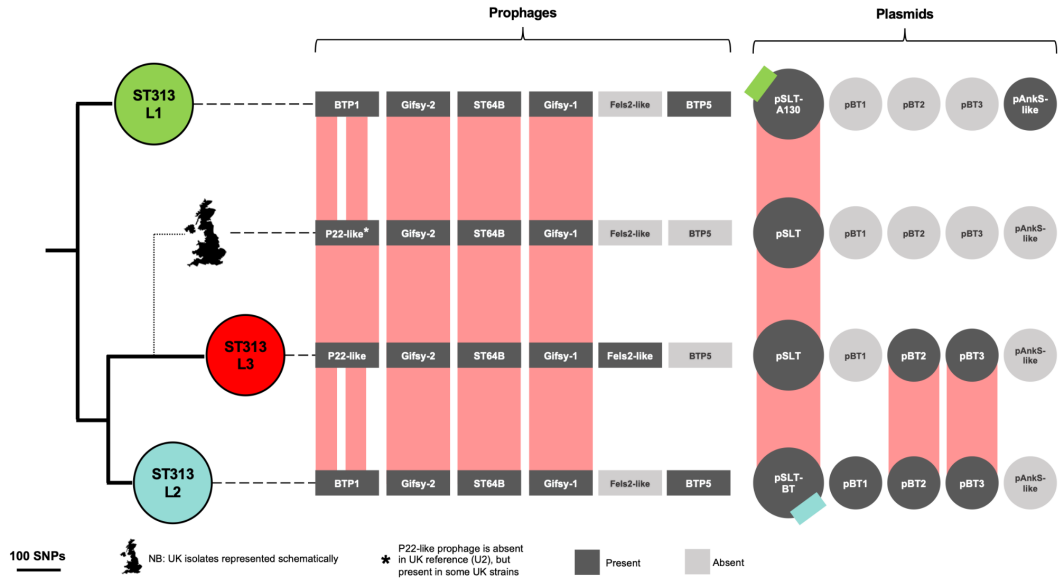


Figure 3.4 The four major ST313 clusters have different prophage and plasmid repertoires

Population structure of ST313 L1, L2 and L3 showing presence/absence of prophages and plasmids using data from **Figure 3.2**. Approximate location of UK-isolated ST313 lineages is indicated schematically due to the diversity of isolates (see **Figure 3.3**). Prophage and plasmid repertoire of lineage reference strains are shown (dark grey indicates presence; light grey represents absence). Reference isolates used; A130 (ST313 L1), D23580 (ST313 L2), BKQZM9 (ST313 L3) and U2 (UK strains). Red blocks represent the extent of conservation between lineage reference genomes, with white gaps indicating missing regions. Coloured squares on the pSLT plasmid represents different AMR cassettes. Note that phylogeny was rooted on ST19, as shown in **Figure 3.3**.

Prophages encode many key *S.* Typhimurium virulence factors (Brussow et al., 2004; Figueroa-Bossi et al., 2001; Ho and Slauch, 2001; Owen et al., 2017; Wagner and Waldor, 2002), and it is clear that different *Salmonella* pathovariants carry distinct prophage repertoires (Wahl et al., 2019). A pairwise comparison of reference strains BKQZM9 (ST313 L3), D23580 (ST313 L2) and U2 (the representative UK-isolated ST313 strain) (**Figure 3.5**) revealed chromosomal differences involving the prophage repertoires, which were generally conserved within lineages (**Figures 2, S1**). The common *S.* Typhimurium prophages Gifsy-1, Gifsy-2 (Figueroa-Bossi et al., 2001) and ST64B (Tucker and Heuzenroeder, 2004) were identified in all three ST313 African lineage reference strains. We found that ST313 L3 carries a Fels-2-like prophage which shares 99.99 % sequence identity to the RE-2010 prophage (Hanna et al., 2012), also found in other extraintestinal *Salmonella* serovars including the global outlier cluster of *S.* Enteritidis (Feasey et al., 2016) and *S.* Panama (**Section 4.3.7**), but absent from the ST313 L2 and the UK-isolated ST313 representatives (**Figure 3.5**).

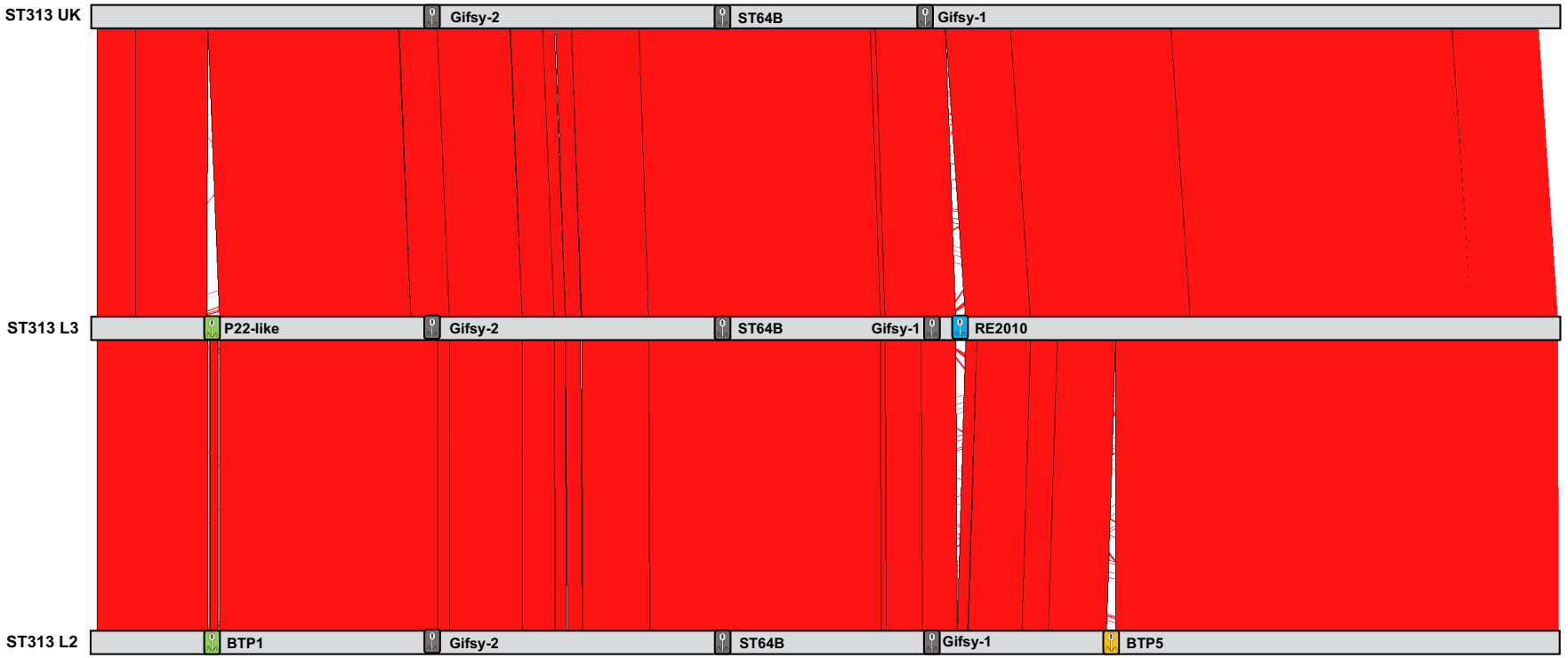


Figure 3.5 Chromosomal comparison of ST313 lineages

Comparison of the complete genome sequences of ST313 lineage representatives generated using ACT (Carver et al., 2005) (see **Section 3.3.14**). Reference isolates used; D23580 (ST313 L2), BKQZM9 (ST313 L3) and U2 (UK-isolated strains). Red represents sequence similarity and white represents regions absent. Prophage positions are represented as coloured boxes.

Prophage BTP5 was absent from the ST313 L3 reference strain. BTP5 is associated with ST313 L1 and L2 (Owen et al., 2017), and generally absent from ST313 isolated in the UK (Ashton et al., 2015). Although prophage BTP1 was not found in ST313 L3, a related P22-like prophage that had 74 % fragmented sequence identity to BTP1 occupied the same attachment site (**Figure 3.6a**). The P22-like prophage had almost identical replication, lysis and capsid genes to BTP1, differing only in the Ea, immunity, tail and *gtr* regions.

The ST313 L3 P22-like prophage contained the *sieB* gene, which encodes a superinfection exclusion phage immunity protein (Susskind et al., 1974). The SieB proteins of ST313 L3 and P22 were 92% identical, with the two proteins sharing 177/192 amino acids. It is possible that the carriage of a gene that mediates abortive phage infection confers a selective advantage to ST313 L3, reflecting similar findings in ST313 L2 (Owen et al., 2020). Specifically, the novel BstA family of prophage defence proteins has recently been identified in the BTP1 prophage of ST313 L2. BstA defends bacterial cells that carry BTP1 against exogenous attack by a variety of lytic phage, without sacrificing lytic autonomy, thus conferring a beneficial trait to ST313 L2 (Owen et al., 2020). Although BTP1 is missing from the UK reference strain U2, two UK-isolated ST313 strains (U15 and U8 (Ashton et al., 2017)) carried a P22-like phage with 99.96 % sequence identity to that found in ST313 L3 (**Figure 3.6b**). In summary, the prophage repertoire of ST313 L3 is more similar to that of the UK-isolated ST313 strains than to African ST313 L1 and L2.

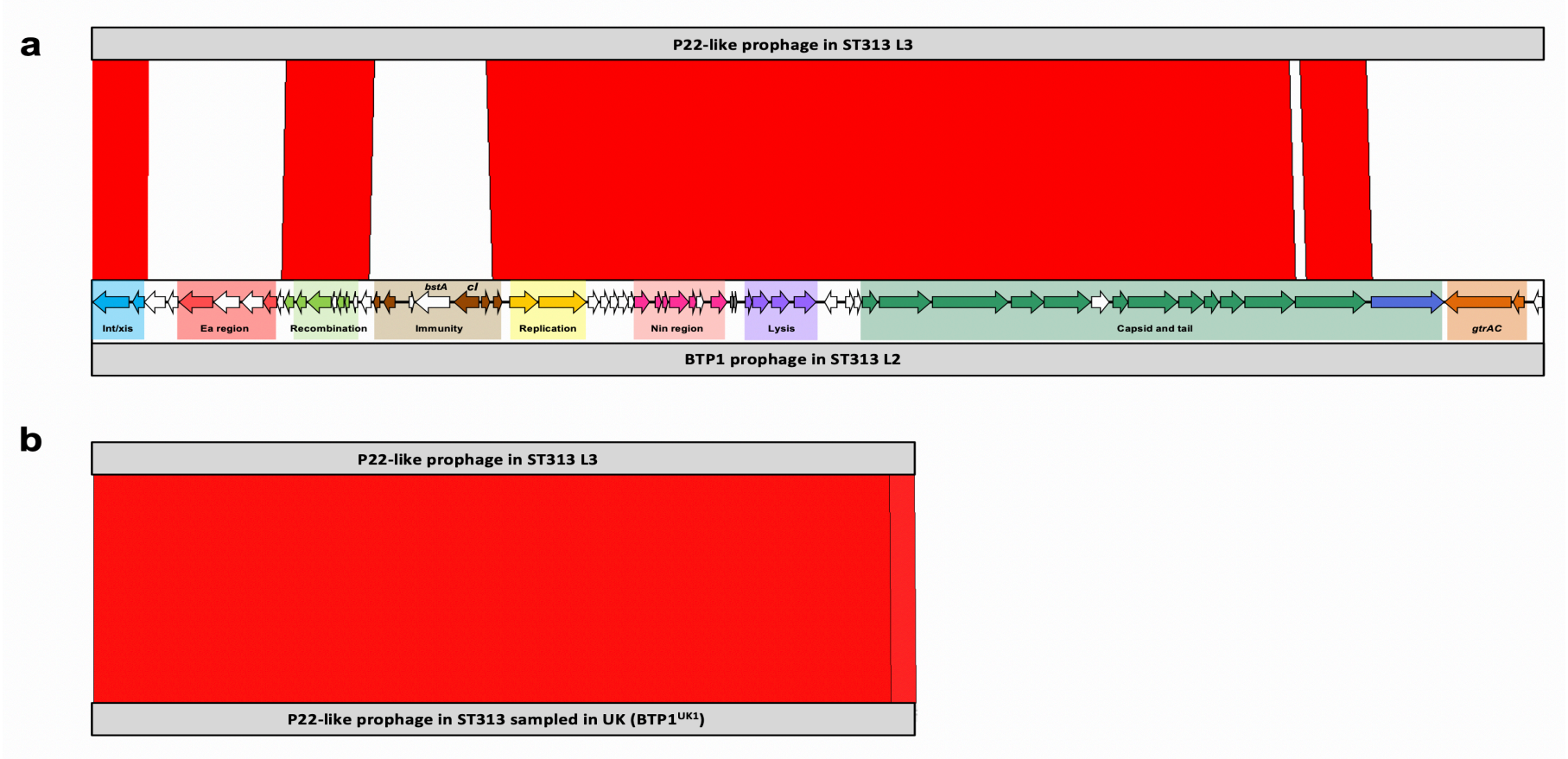


Figure 3.6 Conservation of BTP1 and P22-like prophages in *S.* Typhimurium ST313 L3

Comparison of P22-like prophage regions identified in ST313 lineages generated using ACT (Carver et al., 2005) (see **Section 3.3.14**). Red represents sequence similarity and white represents regions absent. **(a)** shows the comparison between P22-like prophage in ST313 L3 (strain BKQZM9) and ST313 L2 (strain D23580). Prophage annotation is adapted from that shown in Owen *et al.,* 2017 (Owen et al., 2017), with different colours highlighting different prophage regions. **(b)** shows the comparison between P22-like prophage in ST313 L3 (strain BKQZM9) and a P22-like prophage in ST313 sampled in the UK.

3.4.4 Genomic epidemiology of antimicrobial resistance

To obtain epidemiological insights into the emergence of *S.* Typhimurium lineages which cause systemic disease in Africa, the dynamics of AMR were investigated. The draft genome sequences of all 680 isolates were examined for genes and mutations that confer reduced susceptibility to antimicrobials. In total, 65 % (*n* = 440) of isolates had an MDR genotype (Magiorakos et al., 2012). To confirm genome-based predictions, AMR profiles were confirmed phenotypically. The sensitivity and specificity of genome predictions was calculated for 528 available isolates (measured by EUCAST breakpoint). For chloramphenicol; sensitivity was 96.71 % (95 % Confidence Interval [CI] = 94.53 % to 98.19 %) and specificity was 83.60 % (95 % CI = 78.42 % to 87.97 %). For ampicillin; sensitivity was 98.04 % (95 % CI = 96.52 % to 99.02 %) and specificity was 95.61 % (95 % CI = 90.06 % to 98.56 %). For cotrimoxazole; sensitivity was 97.38 % (95 % CI = 95.56 % to 98.60 %) and specificity was 77.09 % (95 % CI = 70.24 % to 83.03 %). Taken together, the data reiterates previous findings in *S.* Typhimurium that show genome-based analysis accurately predicted the AMR phenotypes of 89.8 % isolates with 83% sensitivity and 96% specificity (Mensah et al., 2019).

To determine whether signature AMR profiles accompanied switches in the circulation of ST313 lineages, the genetic determinants of resistance were correlated with lineages (**Figure 3.7**). Amongst isolates belonging to ST19, only 11 % (7/63) were resistant to ≥ 1 antimicrobial. However, of the ST313 lineage isolates, 94 % were resistant to ≥ 1 antimicrobial. There has been a reduction in the proportion of chloramphenicol-resistant isolates of ST313 L2, with 89.19 % (*n*=66/74) prior to 2005, decreasing to 84.11 % (*n*=307/365) between 2006-2015 and 65.28 % (*n*=47/72) between 2016-2018 (**Figure 3.7** and **Table S3.1**). Interestingly, antimicrobial usage policy changes have occurred at the local level in Malawi during this timeframe including the phased withdrawal of chloramphenicol from clinical practice.

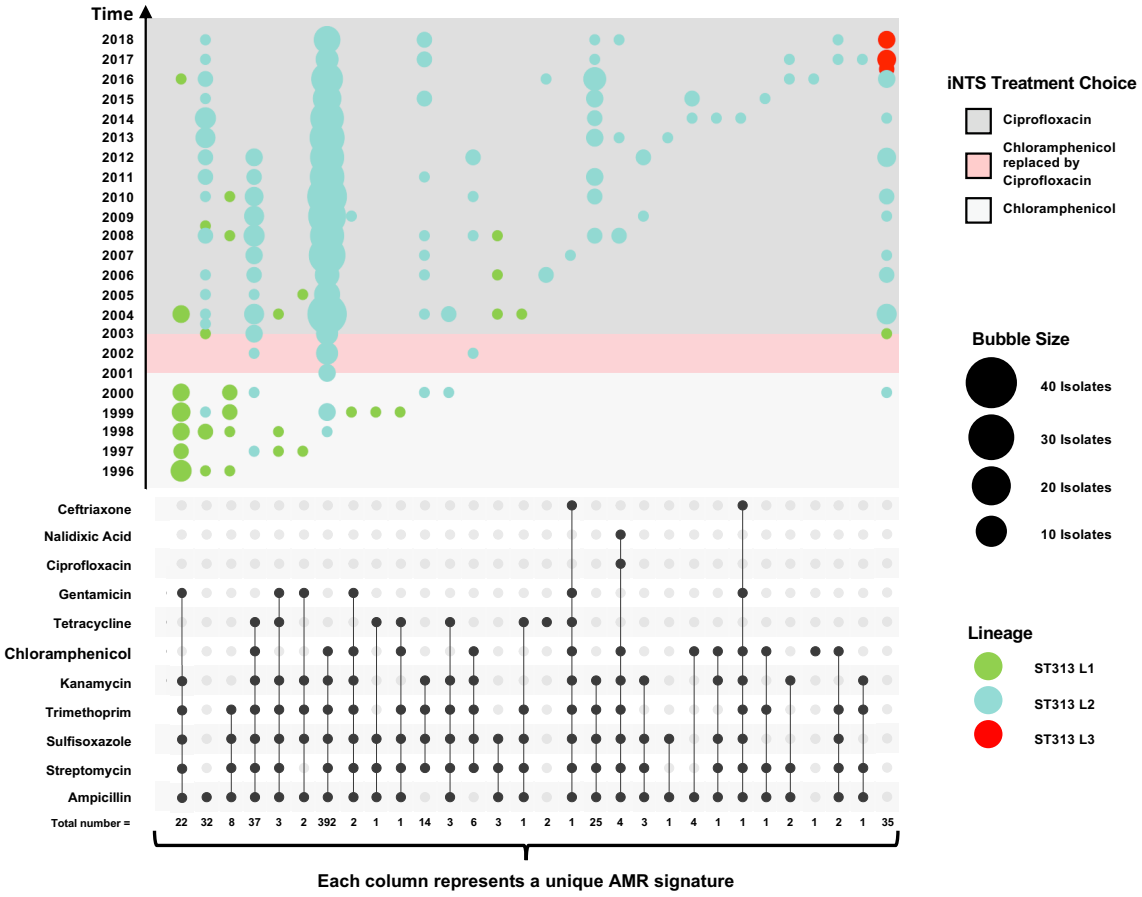


Figure 3.7 Temporal AMR Trends in *S.* Typhimurium lineages (1996 – 2018)

Combination matrix (lower) depicts predicted AMR patterns of *S.* Typhimurium (methods in **Section 3.3.10**). Isolates collected before 1996 were only sampled sporadically, and thus were excluded from analysis. Within the combination matrix, dark grey circles indicate genome-predicted resistance and vertical combination of grey circles represents resistance profile. Total number of isolates with each resistance profile is indicated below matrix. The combination matrix was created using the UpSetR package (Conway et al., 2017). Bubble plot (above) depicts the relative number of isolates (bubble size) with each resistance profile (combination matrix) per year (y axis). Lineage assignments are shown using bubble colour to identify lineage-specific AMR trends. Local Malawi antimicrobial usage policy (Feasey et al., 2015) is highlighted in background colours overlying the timeline in the bubble plot. Bubble plot was created using R ggplot2 (Wickham, 2016).

Overall, the ST313 isolates in this study carried numerous genetic determinants of antibiotic resistance, with 29 different predicted AMR genotype patterns involving 11 antimicrobial agents (**Figure 3.7**). The AMR phenotype of ST313 L1 and L2 was encoded by resistance cassettes carried on a composite Tn*21*-like transposable element, inserted in the *S.* Typhimurium virulence plasmid pSLT (Ashton et al., 2017; Kingsley et al., 2009; Okoro et al., 2015, 2012). The plasmid backbone of pSLT was present in all ST313 strains and lineage-specific variation was observed in the location of the Tn*21* insertion site, consistent with previous reports (Kingsley et al., 2009). The Tn*21*-like region varied between isolates, and reflected the AMR profile (**Figure 3.8**). The plasmid pAnkS$^{ST313\text{-}L1}$ was also implicated in resistance against beta-lactams of ST313 L1, due to carriage of gene *bla*$_{TEM}$ (Sahin et al., 2008).

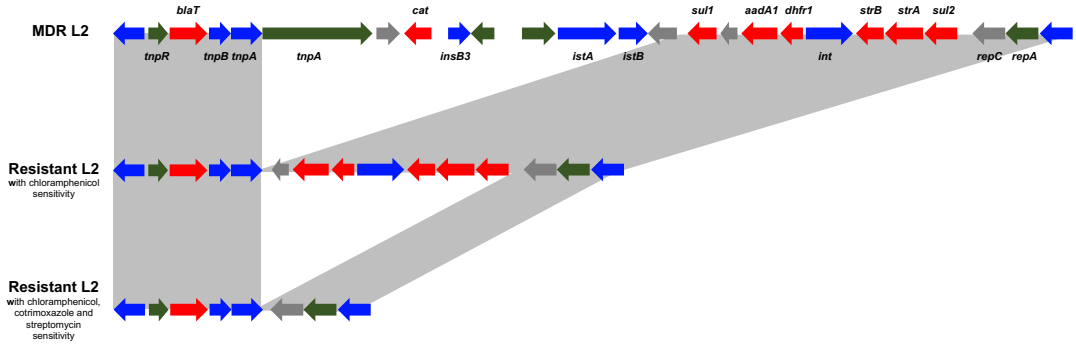


Figure 3.8 Examples of variation within the Tn21-like element

Annotation of the resistance cassette carried on the Tn*21*-like element integrated into the *Salmonella* virulence plasmid pSLT-BT. Examples of variation in the Tn*21*-like element is shown for three different resistance profiles in ST313 L2. MDR L2 refers to isolates with the typical L2 resistance profile of streptomycin, kanamycin, ampicillin, chloramphenicol, trimethoprim and sulfisoxazole. Two additional resistance profiles are shown for L2 as indicated in the figure. Grey boxes between annotations represents gene presence. Annotation is adapted from Kingsley *et al.,* 2009 (Kingsley et al., 2009), and shows antibiotic resistance genes (red), integrase or transposase (blue), pseudogenes (green) and other genes (grey).

ST313 L3 was pan-susceptible to all antimicrobials tested, both genotypically and phenotypically (**Figure 3.3**), as were the UK-isolated ST313 (Ashton et al., 2017). In contrast to L1 and L2, ST313 L3 did not carry an MDR cassette on the pSLT-BT plasmid (**Figure 3.9**).

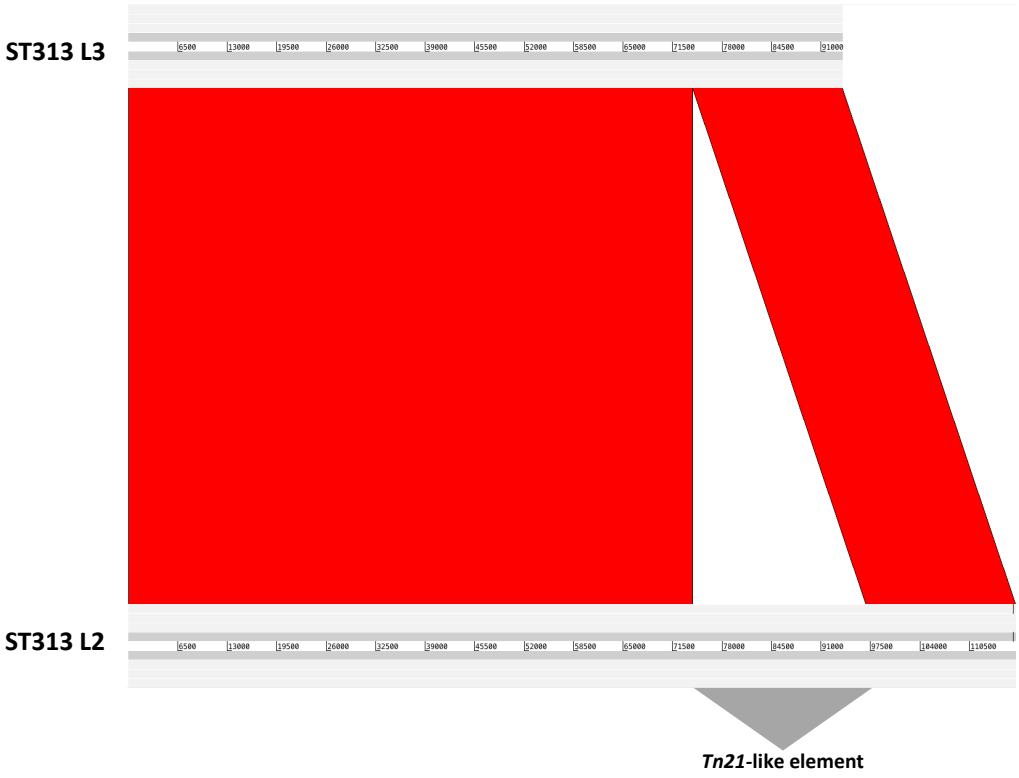


Figure 3.9 pSLT comparison of D23580 vs BKQZM9

Comparison of the *Salmonella* virulence plasmid pSLT identified in ST313 L3 (BKQZM9) and ST313 L2 (D23580) generated using ACT (Carver et al., 2005) (see **Section 3.3.14**). Red represents sequence similarity and white represents regions absent. The location of the Tn*21*-like element is indicated.

Phylogenetic reconstruction of the pSLT plasmid (**Figure 3.10**) reflected the core genome SNP-based phylogeny, demonstrating four individual clades (ST313 L1, L2, L3 and ST19). No evidence of scars associated with plasmid excision at the usual insertion site of the Tn*21*-like transposable element on the pSLT plasmid were found in ST313 L3. Taken together, these findings are consistent with the pSLT plasmid of ST313 L3 never having harboured a Tn*21* transposon.

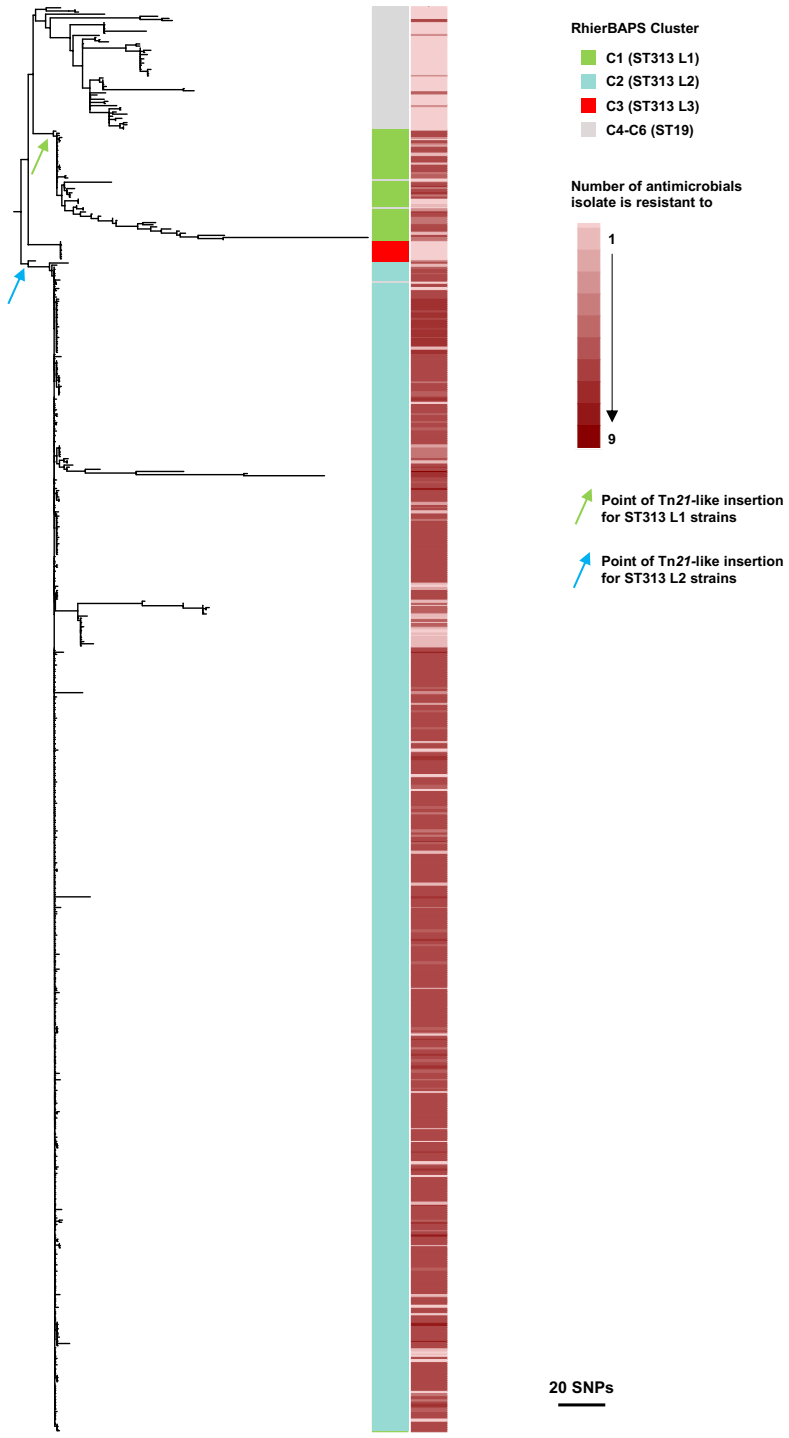


Figure 3.10 Phylogenetic reconstruction of pSLT plasmid

Maximum likelihood phylogenetic tree showing the population structure of the pSLT plasmid in ST19 and ST313 lineages (**Section 3.3.12**). Colour strip from right to left; cluster assignment (lineage) of each isolate and number of antimicrobials each isolate is resistant to. The likely point of insertion of the Tn*21*-like element is indicated for ST313 L1 strains (green arrow) and ST313 L2 strains (blue arrow).

Fluoroquinolone-resistant *S.* Typhi are causing significant clinical challenges, particularly in Asia (Parry et al., 2019). Resistance to fluoroquinolones in high clinical use, such as ciprofloxacin result from a combination of mutations which affect the gene *gyrA*, among others (Hooper and Jacoby, 2015). In total, < 1 % (*n*=5/180) of African *S.* Typhimurium carried mutations associated with reduced susceptibility to fluoroquinolones. Those strains belonged to the phylogenetic lineage ST313 L2 and were isolated in either Malawi (*n*=3), Mali (*n*=1) or Cameroon (*n*=1). The oldest isolate originated from Cameroon in 1998 and carried a single *gyrA* mutation (S83F). Interestingly, this isolate was sequence-typed as ST19 despite belonging to ST313 L2. The Mali isolate, dated from 2008, also carried the same *gyrA* mutation. The three Malawi isolates carried either *gyrA* (D87N) or *gyrA* (D87Y) and were isolated in 2008, 2013 and 2018, after ciprofloxacin was introduced to treat iNTS disease. All five strains were phenotypically susceptible to ciprofloxacin according to results measured by EUCAST breakpoints, consistent with having only a single *gyrA* mutations (Weigel et al., 1998). Ongoing surveillance will be important to detecting any rises in ciprofloxacin resistance, particularly to identify strains with triple mutations which would confer full resistance to this clinically important antibiotic.

3.4.5 The evolutionary trajectory of *S*. Typhimurium ST313

To obtain dates for lineage divergence, Bayesian phylogenetic inference was used (**Figure 3.11, Figure 3.12**). To account for variations in sampling depth in Malawi compared with countries sampled in the contextual collection, and to reduce computation time, I selected 150/680 isolated for BEAST analysis (**Section 3.8.8**). To validate the results from this BEAST analysis, multiple BEAST runs including different random subsets of data could be compared to ensure that the study has not been biased by the sub-sampling only a small number of isolates. I included the prior assumptions of a coalescent Bayesian skyline model for population growth, and a relaxed log normal clock rate to account for rate heterogeneity amongst branches (described in **Data S3.1** and **Methods 3.3.8**).

Based on these prior assumptions, the most recent common ancestor (MRCA) of the ST19 and ST313 sampled in this study date back to the year 918 AD (95 % Highest Posterior Distribution (HPD) = 570-1080) and the MRCA of all ST313 sampled in this study (**Figure 3.12** event 0) was 1566 (95 % HPD = 1271-1953). The dataset used here spanned 50 years and

included many isolates which predate those used in previous phylodynamic analysis (Ashton et al., 2017; Okoro et al., 2012). The incorporation of these historical, more diverse isolates generates a mean value for the MRCA of ST313 lineages that is older than previously described (~1787) (Ashton et al., 2017). However, it should be noted that the branch linked to the MRCA is deeply rooted in the phylogeny and the 95% HPD values reported here have a substantial breadth. A limitation of Bayesian analysis is that it relies upon a series of user-defined prior probability distributions which influence the estimation of tip dates.

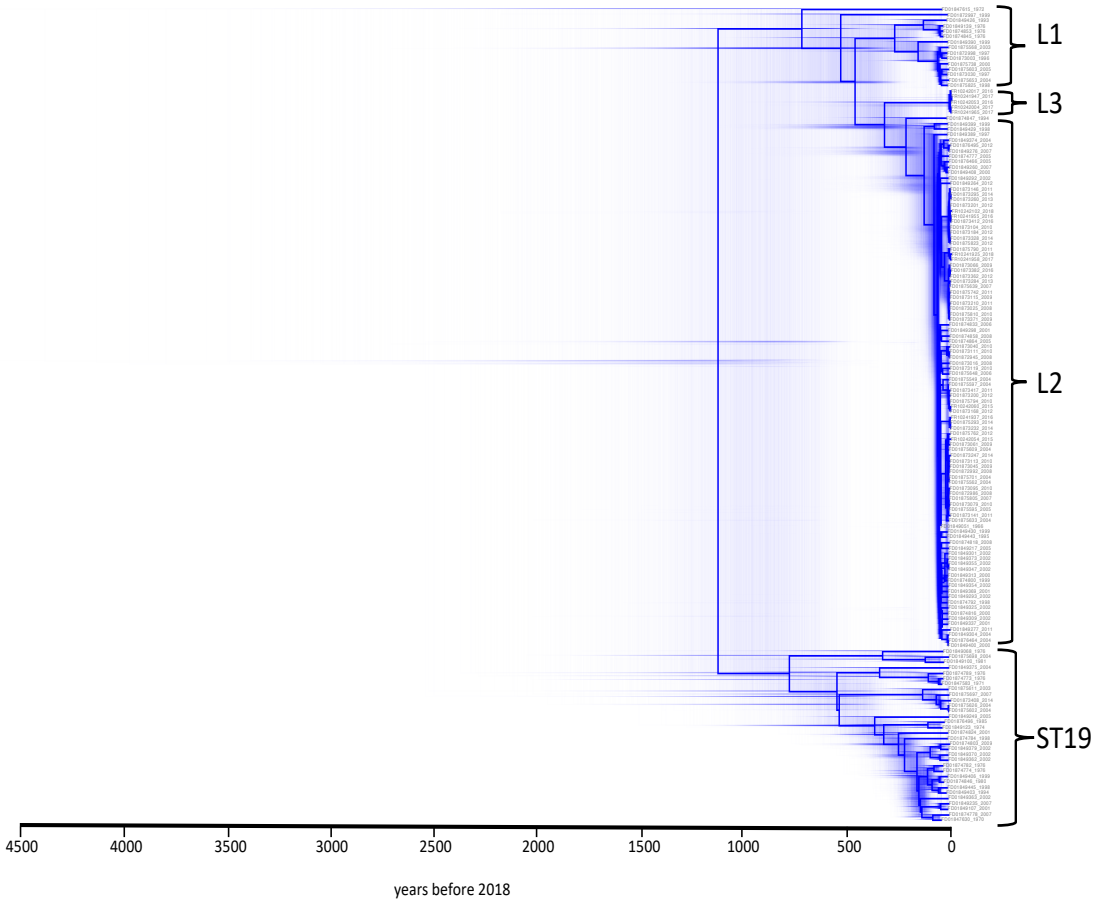


Figure 3.11 Distribution of sampled trees from Bayesian phylogenetic inference

Chronograph of 150 *S.* Typhimurium strains isolated from bloodstream of human iNTS disease patients. Methods are described in **Section 3.3.8**. The figure displays the distribution of possible hierarchies and highlights uncertainty in the timeline. Figure was visualised using DensiTree (Bouckaert, 2010).

The MRCA of ST313 L1 dated to 1794 (95 % HPD = 1738 - 1965), with Malawian ST313 L1 forming a discrete sub-lineage with an MRCA (**Figure 3.12** event 1) dated to 1950 (95 % HPD = 1921-1986). My estimation for the MRCA of ST313 L1 in Malawi confirms previously reported estimates of ~1960 (95 % HPD = 1920-1979) (Okoro et al., 2012), with overlapping 95 % HPD values. The ST313 L2 MRCA (**Figure 3.12** event 2) dated to 1948 (95 % HPD = 1929-1959). One 1966 sample predates the original prediction for the emergence of ST313 L2 of ~1977 (95 % HPD = 1957-1968) (Okoro et al., 2012) which may account for my slightly earlier average estimation of MRCA. The 1966 sample represents the first ST313 L2 isolate to have been reported from Africa. The isolate (STM 252 66) originated from a patient in Senegal, and was predicted to be MDR. Although my average estimation for ST313 L2 is slightly earlier than that previously reported, the 95 % HPD values do overlap. I therefore conclude that my findings support previous estimations. The MRCA of ST313 L3 (**Figure 3.12,** event 3) was 2007 (95 % HPD 1998-2012), showing that the clonal expansion of ST313 L3 occurred in recent evolutionary history. Overall, this analysis has advanced the current understanding of landmark dates in the evolutionary history of ST313 lineages by including a mixture of contemporary and historical ST313 in the dataset.

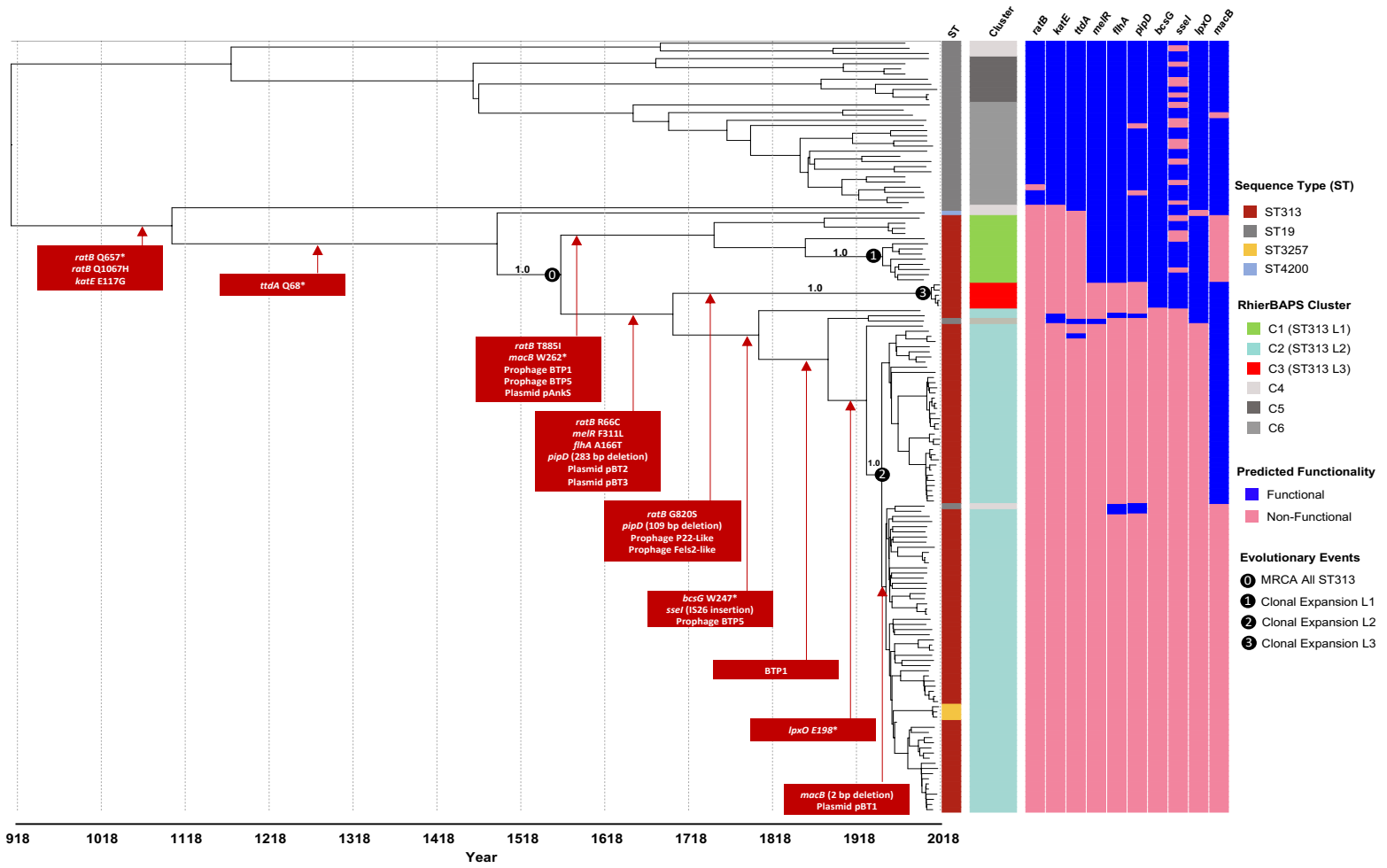


Figure 3.12 Stepwise evolution of *S.* Typhimurium responsible for bloodstream infections in Africa

Legend on next page

Figure 3.12 Stepwise evolution of *S.* Typhimurium responsible for bloodstream infections in Africa

Chronograph of 150/680 *S.* Typhimurium strains isolated from bloodstream of human iNTS disease patients (**Section 3.3.8**). The figure displays a maximum clade credibility tree. Posterior support is shown for key branches, all of which had the maximum posterior probability of 1.0. Adjacent colour strips are as follows (from left to right); ST, lineage assignment (rHierBAPs) with the three major ST313 lineages highlighted in colour. Predicted functionality (see methods) is depicted as a colour strip for each gene and is based on whole genome-based predictions of SNPs likely to play a functional role. Genome degradation events that generate functionally relevant pseudogenes are displayed in red boxes overlying the chronograph. An asterisk (*) represents a premature stop codon. Numbers indicate four key evolutionary events. Figure visualised using iTOL (Letunic and Bork, 2016).

3.4.6 The phylodynamics of pseudogene formation in of *S.* Typhimurium ST313

To explore the known association of pseudogenisation and invasive salmonellosis in my dataset, I put pseudogene formation into a phylogenetic-temporal context. The accumulation of pseudogenes is a feature of invasive, host-adapted bacterial pathogens such as *S.* Typhi and *S.* Paratyphi (Holt et al., 2009) and has been recognised as a signature of adaptation in ST313, termed genome degradation (Kingsley et al., 2009; Okoro et al., 2015). For example, a common theme of the loss-of-function of certain metabolic pathways has been discovered in invasive pathogens (Wheeler et al., 2018). A list of previously identified pseudogenes associated with the emergence of ST313 lineages in Africa was compiled (**Table S3.4**) (Okoro et al., 2015). The collection included metabolic genes (*melR* and *ttdA*), environmental response genes (*katE, macB* and *bcsG*), structurally relevant genes (*ratB, flhA* and *lpxO*) and virulence genes (*sseI, pipD*), and excluded genes of unknown function (Bogomolnaya et al., 2013; Canals et al., 2019; Carden et al., 2017; Gibbons et al., 2000; Lawley et al., 2006; Okoro et al., 2015; Singletary et al., 2016). The SNPs responsible for pseudogene formation were detected, as well as other novel non-synonymous mutations. The conservation of pseudogene-associated mutations was identified at the population level and integrated into the temporal phylogenetic reconstruction. By measuring the evolutionary model of ST313 lineages, the stepwise accumulation of pseudogenes was characterised and hence the progression of functional gene loss associated with host adaptation that likely facilitated the ST313 epidemic in Africa. This phenomenon is exemplified by the timeline of pseudogene accumulation in ST313 L2 shown in **Figure 3.12**.

Recently, the functional significance of the pseudogenisation of *macB* for ST313 L2 was reported (Honeycutt et al., 2020). Here, I report the frameshift event that made the *macB* gene non-functional occurred between 1948 and 1953 (95% HPD = 1931-1963) and represents the final stage in the evolution of ST313 L2 (**Figure 3.12**).

ST313 L3 can be considered as a phylogenetic intermediate between ST313 L1 and ST313 L2. Here, the term "phylogenetic intermediate" is used to reflect the divergence pattern observed in my Bayesian chronogram (**Figure 3.12**). Specifically, ST313 and ST19 diverged first, followed by ST313 L1 and ST313 L3/L2, and finally ST313 L3 and L2. Furthermore, isolates belonging to ST313 L3 contain several pseudogenes which are not present in L1, but are present in L2. The discovery of ST313 L3 therefore allowed further scrutiny of the timeline of gene degradation events. For example, the mutations in *melR* (F311L)*, flhA* (A166T) and *pipD* (283 bp deletion) occurred prior to the divergence of ST313 L3 and ST313 L2. The mutations in *bscG* (W247*), *sseI* (IS*26* insertion) and *lpxO* (E198*) preceded the clonal expansion of ST313 L2, but were not found in ST313 L3. A small number of gene degradation events were ST313 L3-associated, specifically an extra non-synonymous mutation in *ratB* (G820S) and an additional 109bp deletion in *pipD.* Both mutations occurred in genes which had been pseudogenised earlier in evolutionary history and so were unlikely to cause specific phenotypic changes. The fact that *ratB* contained four different mutations predicted to impair gene function, and *pipD* carried two independent mutations, is consistent with both genes being negatively selected during the evolution of ST313 L3. The *macB, sseI* and *lpxO* genes that were non-functional in ST313 L2 were functional in ST313 L3, as previously reported for the UK-isolated ST313 representative (strain U2) (Ashton et al., 2017). I conclude that the pseudogenisation of the *macB, sseI* and *lpxO* genes was a landmark evolutionary event for the development of ST313 L2.

3.4.7 Function of evolutionarily significant *S.* Typhimurium ST313 genes

My genome-derived evolutionary insights were combined with phenotypic experiments that were performed with clinical *S.* Typhimurium isolates with particular genetic differences in specific genes. These experiments allowed the number of lineage-specific phenotypes to be linked to individual SNPs (detailed fully in **Table S3.4**). For example, the KatE catalase protects bacteria from oxidative stress in the environment, and has a high level of activity in stationary phase cultures of *S.* Typhimurium ST19 (Singletary et al., 2016). It was reported previously that the E117G KatE mutation was responsible for a low level of catalase activity in ST313 L1, ST313 L2, ST313 L2.1 and UK-isolated ST313 (Ashton et al., 2017; Singletary et al., 2016; Van Puyvelde et al., 2019), and here it was determined that ST313 L3 had a similarly low level of activity (determined experimentally by X. Zhu, University of Liverpool, UK) (**Table S3.4**). Catalase is required for multicellular growth as biofilms (Ma and Eaton, 1992), a

phenotype associated with survival outside the mammalian host, leading to the suggestion that pseudogenisation of *katE* in ST313 lineages reflects adaptation to a more restricted host-range.

The RDAR-negative phenotype of ST313 L2 (MacKenzie et al., 2019) is caused by the pseudogenisation of *bcsG* (Singletary et al., 2016), a gene that encodes a cellulose biosynthetic enzyme required for biofilm formation. ST313 L3 has an intermediate RDAR phenotype, with an incomplete wrinkling pattern on the colony surface (experimentally determined by B. Perez Sepulveda, University of Liverpool, UK). As the *bcsG* gene is functional in ST313 L3, the genetic basis of the intermediate RDAR phenotype of ST313 L3 is unknown.

Taken together, the loss of traits required for stress-resistance and biofilm formation by ST313 L2 and ST313 L3 is consistent with a reduced requirement for environmental survival by these lineages, and provides additional evidence that ST313 is spread by human-human transmission (Crump et al., 2020; Post et al., 2019).

Adaptation to an invasive lifestyle by ST313 is associated with a decreased ability to colonise and survive within the gastrointestinal tract (Kingsley et al., 2009; Okoro et al., 2015). *S.* Typhimurium ST19 relies upon carbon metabolism to colonise the mammalian gastrointestinal tract (Nguyen et al., 2020) whereas genes required for utilisation of specific carbon sources, such as tartrate and melibiose, are not functional in ST313 lineages (Canals et al., 2019). ST313 L3 was unable to grow with tartrate or melibiose as a sole carbon source (shown experimentally by B. Perez-Sepulveda and H. Webster, University of Liverpool, UK), consistent with the pseudogenisation of the *ttdA* and *melR* genes in both ST313 L3 and ST313 L2 (Kingsley et al., 2009), (Canals et al., 2019) (**Figure 3.12**).

Further evidence of genome degradation, consistent with niche adaptation, is provided by the pseudogenisation of the genes *ratB* in all ST313 lineages, *pipD* in ST313 L2 and L3, and *lpxO* in ST313 L2. The PipD effector protein has been implicated in gastrointestinal pathogenesis of ST19, although a causal relationship has not been demonstrated (Lawley et al., 2006; Wood et al., 1998). The *lpxO* gene is pseudogenised by a stop codon in ST313 L2

(Kingsley et al., 2009). LpxO hydroxylates lipid A, a modification required for virulence of S. Typhimurium ST19 (Moreira et al., 2013). R. Canals (University of Liverpool, UK), J. Bengoechea and A. Dumigan (Queen's University, Belfast) used mass spectrometry to show that the lack of functional LpxO caused structural modifications of Lipid A in ST313 L2 (**Table S3.4**). The *lpxO* gene is functional in ST313 L3, reflecting a different evolutionary path for this lineage.

3.4.8 Invasiveness potential of novel ST313 L3

As detailed above, genome degradation is a key feature in the evolution of bacterial pathogens that are adapted to systemic infection. Recently, the "invasiveness index" was defined as a value that represents the extent of genome degradation and diversifying-selection that was specific to invasive serovars, and was based on a proven set of extraintestinal predictor genes (Wheeler et al., 2018). To determine the potential of ST313 L3 to cause extraintestinal disease, the invasiveness index of each sample in this study was calculated and compared between different ST313 lineages using the Wilcoxon Mann Whitney test (**Figure 3.13**). Because the isolates used in this study originated from human bloodstream, I used the invasiveness index model that was pre-trained by Wheeler *et al.,* because it incorporated a mixture of gastrointestinal and extraintestinal salmonellae (Wheeler et al., 2018). Consistent with previous studies, there was a significant increase in the invasiveness index of ST313 L2 compared with ST313 L1 ($W=7636.5$, $p<0.001$), and ST313 L1 compared with ST19 ($W=552.5$, $p<0.001$) (Van Puyvelde et al., 2019; Wheeler et al., 2018). ST313 L3 had a particularly high invasiveness index (median=0.187, standard deviation (SD)=0.008), which was significantly greater than that of ST19 (median=0.110, SD=0.017), ST313 L1 (median=0.129, SD=0.008) or ST313 L2 (median=0.134, SD=0.006) ($W=0$, $p<0.001$) (**Figure 3.13**). ST313 L3 had a significantly greater invasiveness index than UK-isolated ST313 (median=0.134, SD=0.018)($W=480$, p-value<0.001).

The invasiveness index is determined by assessing the level of degradation in 196 top predictor genes of *Salmonella enterica,* ranked in order of variable importance (Wheeler et al., 2018). In total, 17/196 genes had undergone additional degradation in ST313 L3 (isolate BKQZM9), compared with ST313 L2 (D23580), involving the presence of non-synonymous SNPs, indels or gene loss (**Table S3.5**). A number of the degraded genes in ST313 L3 are required for colonisation of the gastrointestinal tract by *S.* Typhimurium, including *mrcB*,

which was ranked highest in terms of importance. The *mrcB* gene, encodes penicillin binding protein 1b (PBP1b), a major target of β-lactam antibiotics and is important for survival in the presence of bile in *Salmonella* Typhi (Langridge et al., 2009b). In *S.* Typhimurium, a functional *damX* gene is required for resistance to bile (Lopez-Garrido et al., 2010), and was degraded in seven of the nine ST313 L3 isolates. The *napA* gene had extensive deletions in all but one ST313 L3 isolate, and encodes a periplasmic nitrate reductase required for gut colonisation (Canals et al., 2019; Lopez et al., 2015). Taken together, these findings suggest that ST313 L3 is in the process of adapting from an intestinal to a systemic lifestyle.

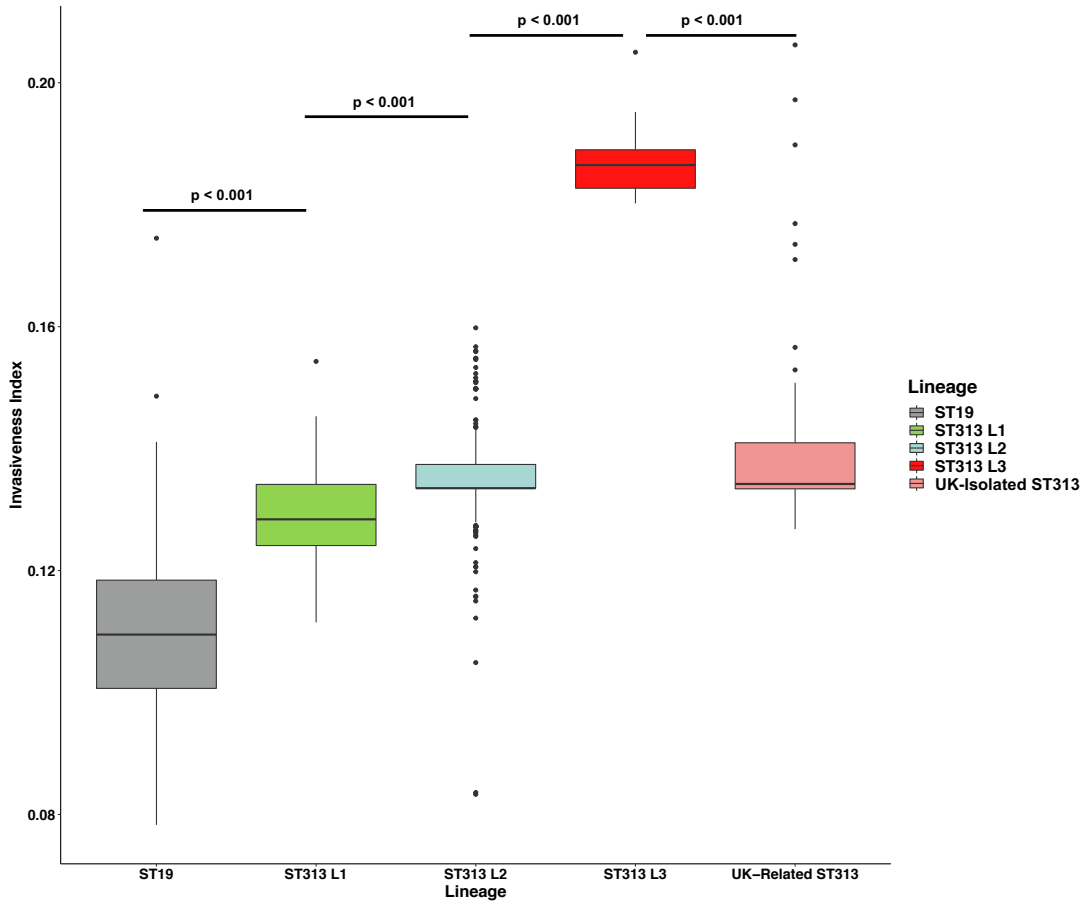


Figure 3.13 Invasiveness index of ST19 and ST313 lineages

Box plot representing the distribution of invasiveness index values for all sequences included in this study summarised by lineage assignment**.** Groups were compared using the Wilcox Mann Whitney test and the resultant *p*-values are displayed above groups. Boxplot centre lines represent median values, boxplot limits represent upper and lower quartiles, boxplot whiskers represent the 1.5 interquartile range and individual points represent outliers. Boxplot was created using R ggplot2 (Wickham, 2016).

3.5 Perspective

Together, my data provide an expanded and contemporary insight into the dynamics of *S.* Typhimurium responsible for systemic disease in Africa, elucidating the stepwise evolution of ST313. This study provides a snapshot of the large repertoire of genomic changes which have shaped the emergence of successful ST313 lineages. Host-adapted *S. enterica* serovars generally cause systemic disease, and often carry pseudogene mutations that prevent the production of certain effector proteins that are dispensable (or disadvantageous) for systemic infections (Klemm and Dougan, 2016). Therefore, genes which promote survival and growth in the gastrointestinal tract are often non-functional in host adapted, extraintestinal *Salmonella* (Klemm et al., 2016; McClelland et al., 2004; Nuccio and Bäumler, 2014). The stepwise pseudogenisation of genes reported here is consistent with the host adaptation of *S*. Typhimurium ST313 lineages over evolutionary history and supports mounting evidence for a human ST313 reservoir (Post et al., 2019; Wheeler et al., 2018). My findings resemble the patterns of evolution as found in other *Salmonella* serovars that currently cause endemics of bloodstream infections globally. For example, *S.* Enteritidis represents a third of all iNTS disease cases in Africa and shows similar signs of genome degradation consistent with niche-adaptation (Reddy et al., 2010).

I discovered that ST313 L3 emerged relatively recently as a cause of systemic disease in the Malawian population, and was first observed in 2016 with an MRCA of ~2007. It is possible that the phased removal of chloramphenicol from clinical practice following local policy changes in Malawi between 2002 and 2005 created a window of opportunity for the emergence of the fully-susceptible ST313 L3.

Despite its contemporary nature, ST313 L3 is an evolutionary intermediate between ST313 L1 and ST313 L2, forming a monophyletic cluster within a diverse group of ST313 originally identified in the UK. These UK-isolated ST313 could also be considered to be an intermediate phylogenetic group, albeit much more diverse that ST313 L3. A limitation of my analysis is that UK-isolated ST313 were not included in the Bayesian temporal phylogeny. Given the placement of ST313 L3 within the UK-isolated ST313 on the phylogeny (**Figure 3.3**), including the UK-isolated strains would facilitate more scrutiny of the evolutionary steps involved in the trajectory towards ST313 L3.

After combining my phylogenetic and accessory genome analysis, I conclude that ST313 L3 has been newly introduced into the Malawian population likely as a result of an international transmission event. Importantly, ST313 L3 had an elevated invasiveness index compared to ST313 L2, making it distinct from UK-isolated ST313 strains which had a lower invasiveness index than ST313 L2 (Ashton et al., 2017). The majority of ST313 L3-unique genes were plasmid or prophage-encoded, including the prophage RE-2010 (Wheeler et al., 2018) which is found in other extraintestinal *Salmonella* pathovariants (Hanna et al., 2012). Aside from accessory genome composition, the majority of ST313 L3-specific changes involve loss-of-function SNPs in genes that are not required for systemic infection. These findings predict that ST313 L3 is better adapted to cause extraintestinal infection than other ST313 lineages from both the UK and in Malawi, a property that could have contributed the emergence of ST313 L3.

Chapter 4

Salmonella Panama: The genomic epidemiology of an understudied serovar causing gastrointestinal infection and invasive disease worldwide

4.1 Acknowledgement of the specific contribution of collaborators

I acknowledge the following contribution of collaborators to the phenotypic experiments described in this chapter. Unless specified below, all work was completed by myself.

Rebecca J. Bengtsson University of Liverpool, UK	Provided bioinformatic support, particularly with GWAS analysis and helped with troubleshooting and narrative direction.
Blanca M. Perez Sepulveda University of Liverpool, UK	Assisted with conceptualisation, WGS and provided expertise on prophage and gene functions.
Rebecca J. Bennett University of Liverpool, UK	Provided bioinformatic support, particularly with GWAS and BEAST analysis and helped with troubleshooting and narrative direction.
Nicolas Wenner University of Liverpool, UK	Assisted in the analysis of prophage sequences
Danielle Ingle Doherty Institute, Australia	Selection of samples from Doherty Institute
George E. Stenhouse University of Liverpool, UK	Provided bioinformatic support, particularly with BEAST analysis and helped with troubleshooting and narrative direction.
P. Malaka De Silva University of Liverpool, UK	Helped with narrative direction and reviewing results
Marie A. Chattaway Public Health England, UK	Provided *S.* Panama samples from Public Health England
Deborah Williamson University of Liverpool, UK	Provided *S.* Panama samples from Doherty Institute
Jay C.D. Hinton University of Liverpool, UK	Supervision and conceptualisation.
François-Xavier Weill Institut Pasteur, France	Supervision, conceptualisation and methodology design.
Kate S. Baker University of Liverpool	Supervision, conceptualisation and methodology design.

4.2 Introduction

NTS disease poses a significant burden to public health globally, causing approximately 153 million cases and 57,000 deaths per annum (Kirk et al., 2015). Over the past decade, just six nontyphoidal *Salmonella* serovars have been responsible for ~70% of cases of human disease in Europe and the USA, namely Enteritidis, Typhimurium, monophasic Typhimurium 1,4,[5],12:i:-, Infantis, Newport and Javiana (Centers for Disease Control and Prevention, 2018; European Food Safety Authority and European Centre for Disease Prevention and Control (EFSA and ECDC), 2019). Of these, only *S.* Typhimurium and *S.* Enteritidis have been investigated intensively, however little research effort has focused on other NTS serovars that cause human disease globally, a disparity that needs to be addressed, particularly for those serovars which cause the most severe disease.

As well as being responsible for gastroenteritis in humans, NTS can also invade normally sterile sites resulting in invasive infections such as bacteraemia and meningitis (Feasey et al., 2012). This iNTS disease was responsible for 535,000 cases and 59,100 deaths globally in 2017 (Stanaway et al., 2019a). In humans, the clinical presentation of NTS disease depends upon a combination of host immune factors and bacterial features that are specific to individual *Salmonella* pathovariants (de Jong et al., 2012; Gilchrist and MacLennan, 2019; Lokken et al., 2016). The high levels of iNTS disease caused by *S.* Typhimurium and *S.* Enteritidis in sub-Saharan Africa have generated a significant research focus (Ao et al., 2015; Stanaway et al., 2019a). However, as happened for NTS disease, other *Salmonella* serovars associated with iNTS infections have received little attention. One of these neglected serovars is *S.* Panama, which has been responsible for numerous geographically-localised outbreaks of NTS infections worldwide and causes more cases of iNTS than most serovars (Pulford et al., 2019).

Whilst iNTS is most commonly seen in sub-Saharan Africa (Ao et al., 2015; Marchello et al., 2019; Stanaway et al., 2019a), *S.* Panama has a very different geographical distribution. The first isolation of *S.* Panama occurred during an investigation of foodborne infection amongst American soldiers stationed in Panama in 1934 (Leeder, 1956). In French territories of America, such as Martinique, Guadeloupe and French Guiana, *S.* Panama is consistently reported as a leading cause of NTS disease (Gay et al., 2014; Guyomard-Rabenirina et al., 2018; Olive et al., 1996; Papa, 1976) and is significantly associated with causing invasive

infections in children (Guyomard-Rabenirina et al., 2018). Although the proportion of NTS cases caused by *S.* Panama is high in these locations, AMR is reported to remain low (Guyomard-Rabenirina et al., 2018; Olive et al., 1996). In contrast, other regions have seen higher levels of AMR in *S.* Panama (Huang et al., 2013; Kempf and Pietzsch, 1977; Lee et al., 2008; Matsushita et al., 2001; Stephan et al., 1977; van Leeuwen et al., 1982).

In Europe, *S.* Panama caused public health concern after spreading through the pork industry and causing multiple hospital outbreaks in the 1970's and 1980's (Cherubin, 1981; Le Minor and Le Minor, 1981; Lee, 1974; Wilkins and Roberts, 1988). *S*. Panama maintained its ranking as one of the top 20 most frequently isolated serovars in the European Union until 2017 (EFSA and ECDC, 2018) and is a frequent cause of invasive disease with 7% of cases causing extraintestinal infection in England (Wilkins and Roberts, 1988). Similarly, *S.* Panama has been an important cause of iNTS in Asia, as the 11th most frequently isolated serovar between 2001 and 2007 (Hendriksen et al., 2011). Asian isolates have been linked with high levels of antimicrobial resistance, with up to 83% of domestic and imported S. Panama isolates in Tokyo reported to be multidrug resistant by the early 2000s (Lee et al., 2008; Matsushita et al., 2001).

WGS-based approaches have been used to characterise multiple NTS serovars that cause invasive disease such as *S.* Typhimurium and *S*. Enteritidis (Feasey et al., 2016; Kingsley et al., 2009; Okoro et al., 2015, 2012; Van Puyvelde et al., 2019). These studies identified signature genomic features that discriminate between the gastrointestinal and extraintestinal lifestyles of *Salmonella* (Klemm and Dougan, 2016; Klemm et al., 2016). Specifically, functions required for escalating growth within the inflamed gut are often lost in a series of gene pseudogenisation events (Nuccio and Bäumler, 2014; Okoro et al., 2015).

In addition to pseudogenisation, other genomic features have contributed to *Salmonella* invasiveness such as SNPs (Hammarlöf et al., 2018) or the presence of plasmids (Lawley et al., 2006), prophages (Brussow et al., 2004; Figueroa-Bossi et al., 2001; Ho and Slauch, 2001; Owen et al., 2017) and key virulence factors (den Bakker et al., 2011). To investigate associations between genotype and phenotype, GWAS have been valuable (Lees and Bentley, 2016). Although such techniques are in their infancy for bacterial pathogens, GWAS have successfully identified the genetic basis of increased virulence in *Listeria*

monocytogenes (Maury et al., 2016) and *Staphylococcus aureus* (Young et al., 2019), resistance-conferring variants in *Mycobacterium tuberculosis* (Desjardins et al., 2016) and host resistance in *Salmonella* Pullorum (Li et al., 2019).

Although *S.* Panama is a globally disseminated serovar associated with iNTS, it has remained an enigma and has never been investigated in detail. In the current era of genomically informed pathogen surveillance, unravelling the molecular epidemiology and evolutionary history of *S.* Panama will provide a vital baseline of understanding for this important serovar going forward. Here, I used whole genome-based approaches to determine the genomic epidemiology of *S.* Panama collected during routine public health surveillance. I characterised six geographically associated *S.* Panama clades and revealed regional trends in AMR. My findings prompted a further phylodynamic investigation which revealed the migratory rates and spatial-temporal evolution of the serovar. Finally, I determined genetic markers of invasiveness between *S.* Panama clades, showing prophage presence and genome degradation.

4.3 Methods

4.3.1 A dataset of 489 *S.* Panama strains from 27 countries collected over 26 years

S. Panama samples were derived from three archived public health collections. Following sampling, whole-genome sequencing and quality control steps (described below), a total of 489 samples were included in the study. A description of the isolates is summarised in **Table 4.1** and is available with all metadata and genome accession numbers in **Table S4.1**.

Table 4.1 Description of 489 *S.* Panama isolates sourced from three collections.

Collection and location details[a]	Nº samples (n=489)	Nº travel associated (n=60)	Nº samples taken from each source site		
			Stool (n=215)	Extraintestinal[b] (n=233)	Unknown (n=41)
Institut Pasteur **Martinique (*n*= 170), French Guiana (*n*=61), Guadeloupe (*n*=56),** France (*n*=2), Thailand (*n*=9), Cambodia (*n*=5), Europe unspecified (*n*=2), Brazil, (*n*=1), Antilles (*n*=1), Chile (*n*=1), Laos (*n*=1), Martinique (*n*=1), Panama (*n*=1), Saint Lucia (*n*=1), Sri Lanka (*n*=1), Tunisia (*n*=1)	315	25	98	217	-
Public Health England **UK (*n*=114)**, Thailand (*n*=13), Mexico (*n*=7), Barbados (*n*=3), Costa Rica (*n*=2), Asia Unspecified (*n*=1), Cameroon (*n*=1), Caribbean unspecified (*n*=1), Germany (*n*=1), Hong Kong (*n*=1), Malta (*n*=1), Mozambique (*n*=1), Spain (*n*=1), Sri Lanka (*n*=1), USA (*n*=1)	149	35	117	16	16
Doherty Institute **Australia (*n*=25)**	25	-	-	-	25

[a]Bold font refers to sampling locations, non- bold font refers to inferred location based on travel data.
[b]The term extraintestinal includes isolates sourced from blood and cerebral spinal fluid

The largest dataset was sourced from the Unité des Bactéries pathogènes entériques, Institut Pasteur, Paris, France which consists of 2,349 *S.* Panama collected from France and French overseas territories between 1993 and 2016. Due to budgetary constraints on the number of samples that could be sent for WGS, the total dataset was sub-sampled. To capture the maximum global context, all isolates associated with travel outside of France and French overseas territories were selected. The reason for focusing on French overseas was because *S.* Panama is reported to be the most common *Salmonella* serotype in these regions (Guyomard-Rabenirina et al., 2018). To encompass invasive *S.* Panama, all cerebral spinal fluid and bloodstream isolates from French overseas territories and a small number of isolates from Metropolitan France were included. To enable a contrast with non-invasive *S.* Panama, one stool isolate per year from each sampling location available was randomly selected using the excel random number generator (RAND function in Excel, Microsoft). The total number of sampled isolates was 439.

To complement the Institute Pasteur dataset, *S.* Panama which had previously been whole genome sequenced during routine surveillance by the National Infection Service, Public

Health England, London, UK (PHE) and the Doherty Institute, Melbourne, Australia were also included. Specifically, the PHE dataset includes whole genome sequence data from 162 routinely sequenced (Ashton et al., 2016) *S.* Panama collected between 2012 and 2019. The Doherty Institute dataset consists of whole genome sequence data from 25 *S.* Panama collected between 2005 and 2019. Doherty Institute isolates were sampled by Danielle Ingle (Doherty Institute, Australia).

Metadata collected during surveillance questionnaires was used to assign an inferred geographical location to each sample. Where a patient reported a recent trip overseas, the travel destination was considered to be the inferred isolate location (Baker et al., 2018a; Ingle et al., 2019). Where travel was not reported, the sampling location was considered to be the location. Locations were grouped based on United Nations classifications. Specifically, locations belonged to one of four geographical regions; Africa, America (including the Caribbean), Asia/Oceania and Europe.

4.3.2 Whole genome sequencing of *S.* Panama isolates from the Institute Pasteur

Isolates from the Institut Pasteur collection have been stored as stab cultures. To resuscitate samples, a sterile loop was used to transfer culture into 10 ml of Tryptic Soy Broth (Neogen, NCM0004A, which was grown overnight at 37°C. To check for potential contaminants, each sample was also streaked onto Drigalski agar plates (BIO RAD, 63834), which were visually inspected following overnight growth at 37°C. If colonies appeared to be homogenous on Drigalski agar, 100 µL of culture was aliquoted into a FluidX 2D Sequencing Tube (FluidX Ltd, UK). Cultures were heat-killed in a 95°C water bath for 20 minutes to produce thermolysates before DNA extraction (MagAttract kit, Qiagen) and were whole-genome sequenced as part of the 10,000 *Salmonella* Genomes Project (Perez-Sepulveda et al., 2020). Illumina Nextera XT DNA Libraries were prepared (Illumina, FC-131-1096) and sequenced (Illumina HiSeq 4000) in multiplex (768) as 150 bp paired-end reads.

4.3.3 Assembly, quality control and annotation of short sequence reads

To prepare genome sequence data for downstream analysis, reads were assembled, subject to quality control and annotated. Adapter content and raw read quality was assessed using FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC v1.0 (http://multiqc.info). Adapter trimming was then performed using Trimmomatic (Bolger

et al., 2014) v0.36 and low-quality regions were trimmed using Seqtk v1.2-r94 (https://github.com/lh3/seqtk) with the trimfq flag. Adapter and quality trimmed reads were re-assessed for quality using FastQC v0.11.5 and MultiQC v1.0. Quality metrics used to pass sequencing reads were as follows: passed basic quality statistics, per base sequence quality, per base N content and adapter content. GC content distributions were visually inspected to determine any isolates with unusual peaks. Only high-quality reads were used in downstream analysis. Unicycler (Wick et al., 2017) v0.3.0 was used to assemble genomes and QUAST (Gurevich et al., 2013) v4.6.3 was used to evaluate assembly quality to standards consistent with EnteroBase (Alikhan et al., 2018). Specifically, N50 > 20 kb, 600 or fewer contiguous sequences, total number of bases between 4Mbp and 5.8Mbp. Prokka (Seemann, 2014) v1.12 was used to annotate the genomes. SISTR (Yoshida et al., 2016) v1.0.2 was used to confirm serovar designations.

4.3.4 Phylogenetic analysis

Trimmed sequencing data was mapped against the *S.* Panama reference genome (CP012346) using Snippy (https://github.com/tseemann/snippy) v4.4.0 with minimum coverage of 4 and base quality of 25. QualiMap (García-Alcalde et al., 2012) v2.0 was used to assess mapping quality and only samples with greater than 10x genome coverage were included in downstream analysis. The mean coverage depth across all isolates was 45.13x. Snippy-core v4.4.0 was used to build a reference-based pseudogenome for each isolate which was passed through snippy-clean v4.4.0 to replace non-ACTG characters with an N. Subsequently, Gubbins (Croucher et al., 2014) v2.2 was used to remove recombinant regions and invariant sites. The resultant multiple sequence alignment of reference based pseudogenomes (13,213 variant sites) was used to infer a maximum likelihood phylogeny using RAxML-ng (Stamatakis, 2014) v0.6.0 with 100 bootstrap replicates to assess support. To assign clusters, rhierBAPs (Tonkin-Hill et al., 2018) was used specifying two cluster levels, 20 initial clusters and infinite extra rounds. Cluster designations were visualised against the phylogenetic tree using iTOL (Letunic and Bork, 2016) v4.2. Pairwise SNP distances were calculated using snp-dists v0.7.0 (https://github.com/tseemann/snp-dists).

4.3.5 Determination of AMR determinants

Genetic determinants for AMR were identified using staramr v0.5.1 (https://github.com/phac-nml/staramr) against the ResFinder (Zankari et al., 2012) and PointFinder (Zankari et al., 2017) databases. Common resistance patterns were identified

and the respective contigs were extracted from the assembled genomes. To understand the genetic context of regions carrying AMR determinants, the contigs were manually inspected using Artemis (Carver et al., 2012) v10.2. The Prokka-based annotations were confirmed and updated using BLASTx (Altschul et al., 1990) v2.10.1 to compare all coding regions against the non-redundant protein sequence database. PlasmidFinder (Carattoli et al., 2014) was used to predict contigs containing plasmids, and the results were compared with contigs containing AMR determinants. Finally, BLASTn (Altschul et al., 1990) v2.10.1 was used to understand the wider context of relevant genetic regions.

4.3.6 Temporal phylogenetic reconstruction

To determine the evolutionary history of *S.* Panama and to model pathogen movement between regions, a chronogram was produced using Bayesian phylogenetic inference. Due to the complexity of the evolutionary model and the availability of computational resources, 50/489 *S.* Panama were sub-sampled for the analysis. The 50 isolates were selected by stratifying the dataset into regions and selecting ~10% of isolates per region using the excel random number generator (RAND function in Excel, Microsoft). A reference-mapped multiple sequence alignment (3,178 variant sites) was created as described above.

The multiple sequence alignment was loaded into BEAUti (Bouckaert et al., 2019) v2.6.1, using the MASCOT (Marginal Approximation of the Structured Coalescent) template (Müller et al., 2018) to implement a structured coalescent approximation to model migration. Dates and region classifications were extracted from the metadata file and used to specify tip dates and locations. The BEAST model test (Bouckaert and Drummond, 2017) was used to integrate over all possible site models. The clock model was specified as relaxed clock log normal in line with previously published Bayesian analysis for *S. enterica* (Ashton et al., 2017; Okoro et al., 2012; Van Puyvelde et al., 2019). TempEst (Rambaut et al., 2016) v1.5.1 was used to estimate the mean clock rate of $3.67e^{-4}$ substitutions per site per year. The estimated clock rate was set as a prior with a log normal distribution. To avoid very small integration steps (and an unfeasibly small probability of observing a phylogenetic tree), the prior distributions on migration rate and population size were adjusted. Specifically, the migration rate was set at 0.1 with an exponential distribution which will prohibit very large migration rates, whilst not discriminating between small and very small values. The prior distribution on the effective population size was set to log normal, to avoid very small values.

BEAST2 (Bouckaert et al., 2019) v2.6.1 was used execute three independent chains of length 200,000,000 with 10% burn in, logging every 1,000 and accounting for invariant sites (**Data S4.1**). Tracer (Rambaut et al., 2018) v1.7.1 was used to assess convergence, with all parameter effective sampling sizes being > 200. LogCombiner (Bouckaert et al., 2019) v2.6.1 was used to combine tree files and DensiTree (Bouckaert, 2010) v2.2.7 was used for initial visualisation. Finally, a maximum clade credibility tree was created using TreeAnnotator (Bouckaert et al., 2019) v2.6.0. The phylogeny was visualised using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) and iTOL (Letunic and Bork, 2016) v4.2. 0

4.3.7 Determination of invasiveness index

The invasiveness index of each isolate was calculated using previously defined methods, using the 196 top predictor genes for measuring invasiveness of *S. enterica* (Wheeler et al., 2018). The same pre-trained model was used to compare the mean invasiveness index for *S.* Panama with the invasiveness index of validation strains previously described (Wheeler et al., 2018). The distribution of invasiveness index values for each cluster were compared using the Wilcoxon Mann Whitney test implemented through R (R Core Team, 2017) v3.4.0. A custom-made database of the top 196 invasiveness predictor genes was then created from the multi-fasta file provided in Wheeler *et al.* (2018). SRST2 (Inouye et al., 2014) v0.2.0 was used to identify variations between the 196 genes in study isolates.

4.3.8 Genome wide association study

A GWAS was run to determine genomic features potentially associated with invasive disease presentation. First, a GWAS was used to investigate the association between invasiveness and accessory genes. Roary (Page et al., 2015) v3.11.0 was used to generate gene presence/absence data and a core gene alignment. SNP-sites (Page et al., 2016) v2.3.3 was used to extract SNPs and the resultant alignment (14,730 variant sites) was used to construct a maximum likelihood phylogeny using RAxML-ng (Stamatakis, 2014) v0.6.0 with 100 bootstrap replicates to assess branch support. Visualisations were made using iTOL (Letunic and Bork, 2016) v4.2 to confirm tree topology. Scoary (Brynildsrud et al., 2016) v1.6.16 was then used to identify accessory genes associated with an invasive phenotype. A Benjamini Hochberg (BH) corrected *p*-value of 0.05 was used to filter significant results (Benjamini and Hochberg, 1995). To support the Scoary predictions, a Cluster of Orthologous Group (COG)-

based analysis was run using the linear fixed effects model implemented through Pyseer (Lees et al., 2018) v1.3.5. To control for potential confounding, inferred locations were included as a covariate. The minimum allele frequency was set at 5 and the maximum number of dimensions was set at 8. Any result labelled as "bad chi sq" or "high bse" were filtered and used a standard significance cut off of 0.05 for *p*-value. Specifically, the likelihood ratio test (lrt) *p*-value was used which accounts for multiple comparisons. The results for the Scoary and Pyseer-COG analysis were compared to identify genes common to both analyses. The genetic context of this subset of genes was evaluated by manually investigating the contigs using Artemis (Carver et al., 2012) v16.0.0.

A second GWAS was run to determine an association between SNPs and invasiveness. BWA mem (Li and Durbin, 2009) v0.7.10-r789 was used to map trimmed sequencing data against the *S.* Panama reference genome. Mapped reads were then cleaned and sorted using the SAMtools suite (H. Li et al., 2009) v1.7. Reads were realigned against the reference using GATK (McKenna et al., 2010) v3.7 by creating targets for realignment (RealignerTargetCreator) and performing realignment (IndelRealigner). Removal of optical duplicates was completed using Picard v2.10.1-SNAPSHOT (https://broadinstitute.github.io/picard/). Sequence variants were called using Bcftools v1.9-80 (http://samtools.github.io/bcftools). Vcfutils was used for variant filtering as part of the SAMtools suite (H. Li et al., 2009) v1.7. specifying 4 as the minimum depth of coverage.

To run the SNP-based GWAS, the linear fixed effects model was used, implemented through Pyseer (Lees et al., 2018) v1.3.5. Inferred location was set as a covariate and a minimum and maximum allele frequency was set of 0.02 and 0.98 respectively. The lrt *p*-value, corrected for population structure, was used to assess significance. The significance threshold was adjusted for multiple comparisons by dividing the desired significance threshold (0.05) by the total number of variants (3,878) to achieve a suggestive *p*-value of $1.29e^{-5}$. A quantile-quantile (qq) plot was created to look for evidence of *p*-value inflation. A Manhattan plot was created using R ggplot2 (Wickham, 2016). The top hits were manually investigated using Artemis (Carver et al., 2012) v16.0.0.

4.3.9 Prophage annotation and characterisation

The online tool, PHASTER (Arndt et al., 2016) was used to characterise the prophage repertoire of the *S.* Panama reference genome (CP012346). The Fels-2-like prophage sequence was extracted and formatted as a custom-made database for analysis in SRST2. SRST2 (Inouye et al., 2014) v0.2.0 was then used to determine percentage coverage against the prophage region in the *S.* Panama reference. Results were manually mapped onto the phylogeny and chronogram using iTOL (Letunic and Bork, 2016) v4.2 which allowed inference of prophage presence and absence in clades. A representative strain from clade 4 and clade 6 were selected and ran through PHASTER to identify the correct contiguous sequence with the prophage. Artemis (Carver et al., 2012) v16.0.0 was used to extract the prophage-containing contig from the clade 4 and clade 6 representative strains. The prophage regions were annotated by cross-referencing results from BLASTx (Altschul et al., 1990), PHASTER (Arndt et al., 2016) and Prokka (Seemann, 2014). ACT (Carver et al., 2005) was used to perform a pairwise comparison of the clade 4 and clade 6 prophages against the following: *S.* Panama reference strain prophage (CP012346, region 1108684-1155312bp), Fels-2 (NC_010463.1), RE-2010 (HM770079) and SopEΦ (FQ312003.1 region 2855616 to 2901979).

4.4 Results and Discussion

4.4.1 Establishing an informative collection of *S.* Panama isolates summary

To create a relevant dataset, the whole genome sequence data of 489 *S.* Panama isolates was assembled from public health collections (**Table 4.1**, **Table S4.1**). The largest part of the dataset included 315 isolates sourced from the Unité des Bactéries pathogènes entériques, Institut Pasteur (IP), Paris, France. The isolates were sampled from the IP archives which hold 2,349 *S.* Panama collected from France and French overseas territories between 1993 and 2016. The strategy for sampling the IP collection was designed to capture both invasive and non-invasive isolates and put them into a global context. Specifically, 439 samples were selected which incorporated: all isolates associated with overseas travel; all blood and cerebral spinal fluid isolates from French overseas territories; a small set of isolates from metropolitan France; and one stool isolate per year from each sample location available. The majority (315/439) of sampled isolates could be resuscitated, sent for whole genome sequencing and passed sequence quality control.

The IP dataset was complemented by 149 isolates collected in England and Wales, between 2012 and 2019, by the National Infection Service, Public Health England (PHE), London, UK. PHE has routinely genome-sequenced every *Salmonella* isolate received by the *Salmonella* reference service since 2015 (Ashton et al., 2016) in addition to a small number of pre-2015 isolates. The resulting genomic data are regularly deposited in publicly available repositories (Ashton et al., 2016). In total, the sequences of 162 *S.* Panama isolates were downloaded from the Sequence read archive, and 149 passed sequence quality control. Finally, 25 *S.* Panama isolates were analysed which were collected by the Doherty Institute (DI), Melbourne, Australia between 2005 and 2019. All 25 isolates passed QC and were included in the final dataset.

4.4.2 Population structure of *S.* Panama isolates

To define population structure of an understudied serovar, the phylogenetic relationship between 489 *S.* Panama genomes was explored (**Figure 4.1**). The sequences were mapped onto the high quality complete reference genome of *S.* Panama, strain ATCC7378 (Yao et al., 2016) (GenBank Accession: CP012346), recombination was removed and SNPs were identified in the remaining sites. The resulting core genome SNP alignment, of length 13,213 SNPs, was used to construct a maximum likelihood phylogenetic tree with 100 bootstraps to

assess support. Cluster designation using rhierBAPs (Tonkin-Hill et al., 2018) revealed seven clusters. Mapping of rhierBAPs designations onto the phylogenetic tree showed that six of the clusters were monophyletic, with greater than 95% bootstrap support. The six monophyletic clusters were defined as clade 1 to 6, specifically C1 (*n* = 56), C2 (*n* = 65), C3 (*n* = 44), C4 (*n* = 66), C5 (*n* = 95) and C6 (*n* = 68). Mean pairwise SNP distances were calculated both within and between clades (**Table 4.2**), which demonstrated a high level of genetic similarity between isolates belonging in C2 (mean pairwise distance =32 SNPs), followed by C1 (39 SNPs), C3 (64 SNPs), C5 (82 SNPs), C6 (88 SNPs) and C4 (122 SNPs).

Table 4.2 Mean pairwise SNP distances between and within *S.* Panama clades

	C1	C2	C3	C4	C5	C6
C1	**39**	169	234	347	272	492
C2	169	**32**	207	347	245	466
C3	234	207	**64**	395	293	514
C4	374	347	395	**122**	318	605
C5	272	245	293	318	**82**	487
C6	492	466	514	605	487	**88**

NB: Mean pairwise SNP distances within clades are highlighted in light grey

The remaining samples in this study belonged to a seventh, polyphyletic, cluster which was returned as a single cluster on rhierBAPs and included miscellaneous isolates that did not belong to any other clade. Inconsistencies between maximum likelihood phylogenetic clades and rHierBAPs clusters may arise due to intrinsic differences in the methodology of the two approaches. Specifically, rHierBAPs will cluster isolates based on SNP similarity by directly integrating over allele frequencies in each group (Tonkin-Hill et al., 2018), whereas maximum likelihood phylogenetics applies an evolutionary model to the sequencing data (Stamatakis, 2014).

To investigate the relationship between clades and geographic region, location data was mapped onto the phylogenetic tree (**Figure 4.1**). When patients reported a recent overseas journey, the travel destination was designated as the inferred location of the isolate (Baker et al., 2018a; Ingle et al., 2019). Locations were grouped into four regions based on geography and United Nations regional definitions, specifically America (including the Caribbean), Europe, Asia/Oceania, and Africa. Taken together, these findings revealed evidence of regional clustering throughout the phylogenetic tree which may be of relevance for the genome-based surveillance of future outbreaks.

Four clades were predominantly composed of American isolates (96.43% of C1, 96.92% of C2, 100% of C3, 100% of C5). Each of the four clades contained multiple isolates from the same country, suggesting localised clonal circulation (**Figure 4.1**). For example, 100% (44/44) of C3 isolates originated in French Guiana. In contrast, isolates from Martinique were largely spread across two clades, specifically making up 91.07% of C1 and 98.95% of C5. A small proportion of C2 isolates also originated in Martinique (12/65), however the majority were from Guadeloupe (49/65). Phylogenetically proximate isolates from Guadeloupe and Martinique were identified, suggesting that inter-island transmission of C2 may be occurring.

The remaining two clades were mainly composed of European (C4) or Asia/Oceania (C6) isolates. In total, 96.97% of isolates in C4 originated in Europe. Whilst the majority of these isolates came from the UK (56/66), C4 also contained isolates from less-well sampled European countries including one from each of Germany, Malta and France. These findings are consistent with European-wide circulation, and so C4 was designated as the "European clade". A proportion (22.06%) of the C6 isolates were also isolated in Europe, however most (76%) originated in Asia and Oceania, with specific locations including Thailand (22/68), Australia (22/68), Cambodia (5/68), Cameroon (1/68), Hong Kong (1/68) and Laos (1/69). To reflect the origin of the majority of the isolates, C6 was termed the "Asia/Oceania clade".

A key limitation of this analysis was the disparity in sampling depth between countries, reflected in the distribution of isolates which did not belong to any of the six clades. The miscellaneous isolates often originated from less well-sampled countries, suggesting that many more *S.* Panama clades remain to be discovered by a broader sampling strategy.

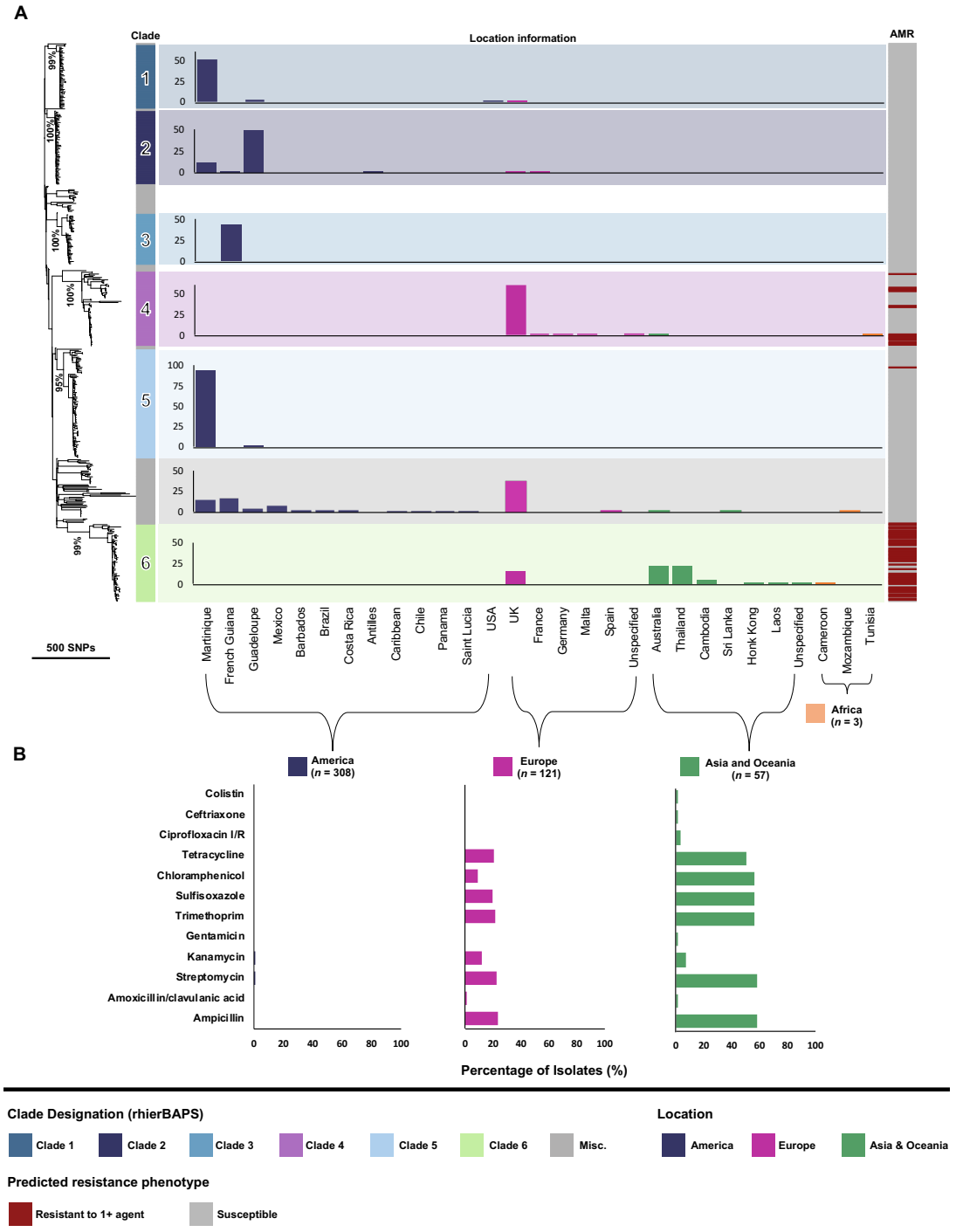


Figure 4.1 Population structure, AMR trends and geographical distribution of *S.* Panama

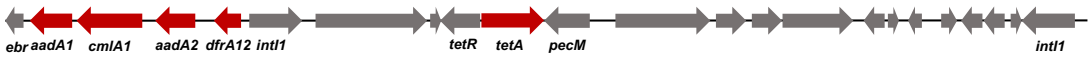
(A) Maximum likelihood phylogeny (**Section 4.3.4**) demonstrating the population structure of 489 *S.* Panama from 27 countries, collected over 26 years (1993-2019) during routine public health surveillance (**Section 4.3.1**). Bootstrap values are displayed on key branches for C1 to C6. Colour strip immediately adjacent to the phylogeny shows cluster assignment (rHierBAPs), with grey representing a miscellaneous polyphyletic cluster. Bar graphs show the number of isolates in each clade from each inferred location. Locations are further grouped into regions, specifically America (blue), Europe (purple), Asia/Oceania (green) and Africa (orange). AMR colour strip indicates strains carrying >1 antimicrobial resistance determinant (red), or strains which are predicted to be pan-susceptible (grey) (**Section 4.3.5**). (B) Horizontal bar charts show the percentage of isolates from each region carrying resistance determinants to each antimicrobial. Note that isolates from Africa were not shown, due to low numbers. Position of root is inferred from Bayesian analysis (**Figure 4.5**).

4.4.3 Antimicrobial resistance trends

To determine AMR trends in this collection of *S.* Panama isolates, the draft genome sequences of all 489 samples were examined for genes and mutations that confer reduced susceptibility to antimicrobials. Previous studies have demonstrated that genome-based AMR analysis accurately predicted phenotype for 97.82% of *Salmonella* isolates (Neuert et al., 2018). In total, 13.91% (68/491) of isolates were predicted to be resistant to at least one antimicrobial class, with the remaining 86.06% (421/489) predicted to be pan-susceptible, lacking any known antibiotic resistance genes or mutations. The majority of resistant isolates (66/68) fell within C4 or C6, and originated in Asia/Oceania (36/68) or Europe (29/68) (**Figure 4.1**). These findings are supported by other phenotypic studies which reported an increase in *S.* Panama isolates from Asia that were resistant to multiple antibiotics, including cotrimoxazole, ampicillin, streptomycin, kanamycin and gentamicin (Lee et al., 2008).

C4 is predominantly composed of European isolates, many of which have acquired MDR determinants. C4 may have arisen following an outbreak that involved the continental spread of *S.* Panama across Europe, associated with the food production industry in the 1970's and 1980's (Cherubin, 1981; Le Minor and Le Minor, 1981; Lee, 1974; Wilkins and Roberts, 1988). It is generally believed that these European *S.* Panama has been exposed to a high antibiotic selective pressure in humans or food animals during their evolution and consequently became resistant to multiple antibiotic agents via acquisition of mobile genetic elements (Avril et al., 1977; Bouanchaud and Chabbert, 1969; Guinee, 1969, 1968; Guinee et al., 1967; Manten et al., 1971; van Leeuwen et al., 1982).

The most common resistance profile, occurring in 51.47% (35/68) of resistant isolates included streptomycin (*aadA1, aadA2*), ampicillin (*blaTEM-1B*), chloramphenicol (*cmlA1*), trimethoprim (*dfrA12*), sulfisoxazole (*sul3*) and tetracycline (*tetA*). All 35 isolates belonged to C6 and a substantial proportion were sampled from patients who specifically reported travel to Asia (16/35). Detailed analysis of the draft genomes showed that most of the resistance genes (*aadA1*, *aadA2*, *cmlA1*, *dfrA12* and *tetA*) were encoded by an MDR cassette located on a single contig (**Figure 4.2**). A BLAST search of the contig revealed 99.79% sequence identity to a 21.2kb region of an 83.2kb plasmid (CP044301.1), which had previously been identified in an *E. coli* isolate from Pork in Cambodia (unpublished).



Gene Name (Ordered)	Functional Annotation
ebr	Quaternary ammonium compound efflux SMR transporter
aadA1	Streptomycin resistance gene
cmlA1	Chloramphenicol resistance gene
aadA2	Streptomycin resistance gene
dfrA12	Trimethoprim resistance gene
intI1	Integron integrase
Unknown	Tn3 family transposase
Unknown	Hypothetical protein
tetR	TetR family transcriptional regulator
tetA	Tetracycline resistance gene
pecM	EamA family transporter
Unknown	Tn3 family transposase
Unknown	Plasmid pRiA4b ORF-3 family protein
Unknown	Transposase
Unknown	Hypothetical protein
Unknown	Phage integrase
Unknown	Hypothetical protein
Unknown	Hypothetical protein
Unknown	DEAD/DEAH box helicase
Unknown	IS5-like element ISKpn26 family transposase
Unknown	IS5-like element ISKpn26 family transposase
Unknown	Hypothetical protein
intI1	Integron integrase

Figure 4.2 Functional annotation of the most common *S.* Panama MDR cassette.

Contiguous sequence containing MDR cassette extracted from 35 isolates of *S.* Panama with resistance against streptomycin (*aadA1, aadA2*), chloramphenicol (*cmlA1*), trimethoprim (*dfrA12*), and tetracycline (*tetA*). Resistance determinants are highlighted in red.

The second most common AMR profile, occurring in 19.12% (13/68) of resistant isolates was against streptomycin (*aph(3')-Ib*), kanamycin (*aph(6)-Id*), ampicillin (*blaTEM-1B*), trimethoprim (*dfrA14*), sulfisoxazole (*sul2*) and tetracycline (*tetA*). The 13 isolates were all part of C4 and were isolated from Europe. All genes were carried by an MDR cassette located on a single contig (**Figure 4.3**). The contig shared 99.98% sequence identity with a 28.3kb region of a 50.9kb *S. enterica* plasmid (CP028173.1), which has been identified in multiple strains including an *S.* Enteritidis isolate from a human in Ghana (Kudirkiene et al., 2018).

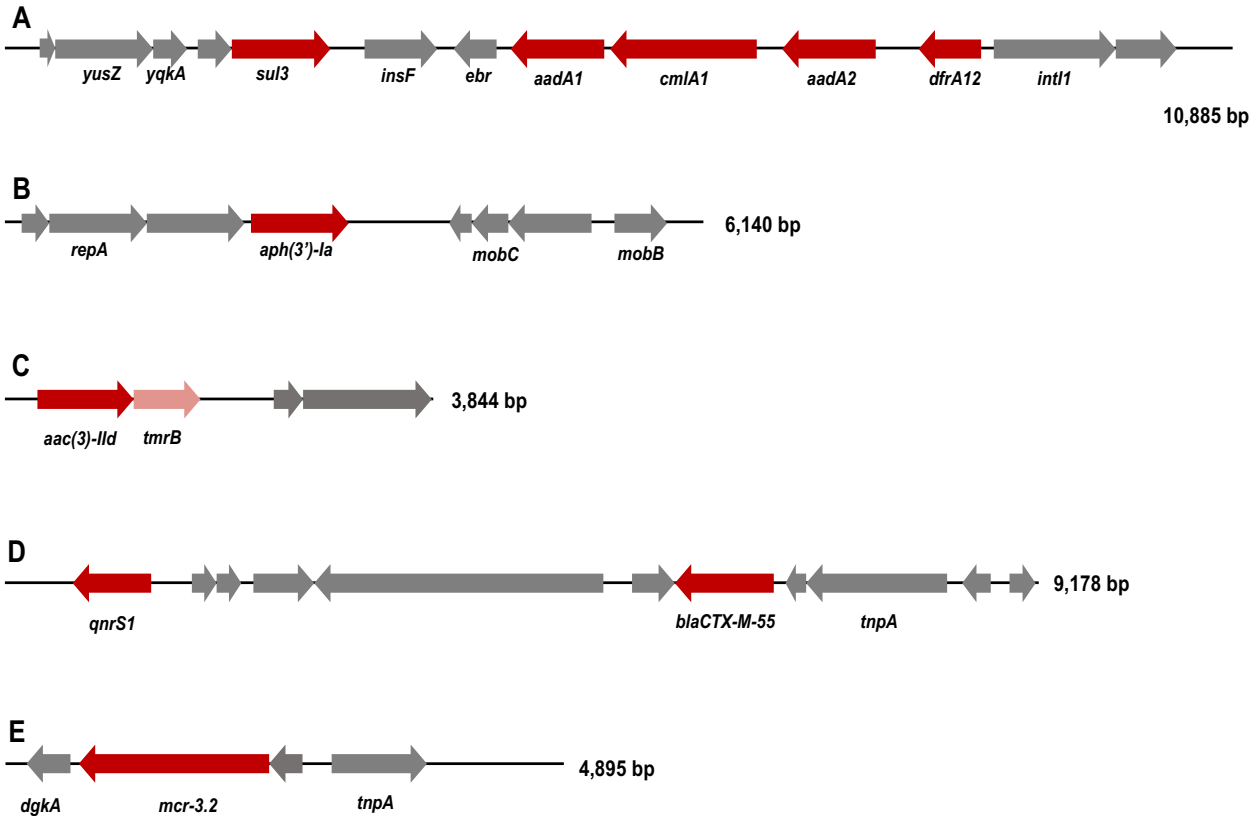
Gene Name (Ordered)	Functional Annotation
Unknown	Anti-restriction protein
Unknown	Hypothetical protein
ccgA2	Hypothetical protein
Unknown	Hypothetical protein
Unknown	Hypothetical protein
Unknown	Transcriptional regulator
Unknown	Hypothetical protein
Unknown	Hypothetical protein
traD	Type IV conjugative transfer system coupling protein
traI	Conjugative transfer relaxase/helicase
Unknown	Hypothetical protein
aph(6)-Id	Kanamycin resistance gene
drfra14	Trimethoprim resistance gene
aph(3')-Ib	Streptomycin resistance gene
sul2	Sulfisoxazole resistance gene
Unknown	Tn3 family transposase
pecM	EamA family transporter
tetA	Tetracycline resistance gene
tetR	TetR family transcriptional regulator
blaTEM-1B	Ampicillin resistance gene
Unknown	Transposase
Unknown	Tn3 family transposase
nuc	Phospholipase D family protein
virB11	P-type DNA transfer ATPase
traI	Conjugal transfer protein
traH	Conjugal transfer protein
traG	Conjugal transfer protein
Unknown	Type IV secretion protein
Unknown	Hypothetical protein
virB5	Type IV secretion system protein
Unknown	Conjugal transfer protein
Unknown	Hypothetical protein
Unknown	Hypothetical protein
Unknown	Hypothetical protein
virB1	Hypothetical protein
hvrA	Trans-acting regulatory protein
Unknown	Hypothetical protein
Unknown	IncN plasmid KikA protein
Unknown	Hypothetical protein
Unknown	Restriction endonuclease
insB	Is1 family transposase
Unknown	DNA cytosine methyltransferase
Unknown	Restriction endonuclease
Unknown	Recombinase family protein

Figure 4.3 Functional annotation of second most common *S.* Panama MDR cassette.

Contiguous sequence containing MDR cassette extracted from 13 *S.* Panama isolates with resistance against streptomycin (*aph(3')-Ib*), kanamycin (*aph(6)-Id*), ampicillin (*blaTEM-1B*), trimethoprim (*dfrA14*), sulfisoxazole (*sul2*) and tetracycline (*tetA*). Resistance determinants are highlighted in red.

The remaining 7.00% of resistant isolates had a variety of susceptibility profiles including one XDR sample which was phylogenetically classified as C6. Epidemiological information showed that the sample was isolated in 2012 from the bloodstream of a 20-year-old male in France who had recently returned from Thailand. The isolate carried AMR determinants specific to eight different antimicrobial classes, including fluoroquinolones, polymyxins and ESBL-mediated resistance to third-generation cephalosporins. In *S.* Typhi, XDR is defined as resistance to five antibiotic agents and includes resistance to fluoroquinolones and ESBL activity. The nomenclature used to describe the extensively drug resistant isolate, is consistent with previous definitions for XDR (Klemm et al., 2018). The resistance genes were distributed across five contigs (**Figure 4.4**). The first contig carried resistance genes against streptomycin (*aadA1, aadA2*), chloramphenicol (*cmlA1*), trimethoprim (*dfrA12*) and sulfisoxazole (*sul3*). A second contig contained resistance genes against ciprofloxacin (*qnrS1*), ampicillin and ceftriaxone (*blaCTX-M-55*). The remaining kanamycin (*aph(3')-Ia*), gentamicin (*aac(3)-IId*) and colistin (*mcr-3.2*) genes were each carried on separate contigs.

These findings of clustered antibiotic resistance elements on the genome could be linked to the plasmid-mediated resistance in *S.* Panama that has previously been reported (Lee et al., 2008). The plasmid complement of the *S*. Panama serovar has rarely been studied in the past, but it is clear that *S.* Panama does not commonly carry the large plasmids (Carneiro et al., 2018), that have previously been associated with virulence in other *Salmonella* serovars (Silva et al., 2017). Rather, *S.* Panama has been reported to carry a heterogeneous population of plasmids (**Table S4.1**) (Stanley et al., 1995). The short read-based plasmid prediction analysis (**Section 4.2.5**) suggested that only 14.72% (72/489) of *S.* Panama isolates in this study carried plasmids. Overall, 76.47% (52/68) of resistant isolates and only 4.75% (20/421) of susceptible isolates were predicted to carry plasmids. The AMR determinants were almost always carried on a different contig to those predicted to be plasmids, which does not necessarily mean that AMR genes are not carried on a plasmid. Specifically, the methodology used to predict plasmid contigs, relies upon the identification of contigs containing replicons identified in the PlasmidFinder database (Carattoli et al., 2014). Due to limitations in short read sequencing data, many plasmid contigs with fragmented assemblies may be missed (Arredondo-Alonso et al., 2017). Long-read sequencing would be required to confirm plasmid presence/absence in individual *S.* Panama isolates.



Gene Name	Functional Annotation (A)
Unknown	Hypothetical protein
yusZ	Multispecies KR domain containing protein
yqkA	Hypothetical protein
Unknown	Hypothetical protein
sul3	Sulfisoxazole resistance gene
insF	IS256 family transposase
ebr	Quaternary ammonium compound efflux SMR transporter
aadA1	Streptomycin resistance gene
cmlA1	Chloramphenicol resistance gene
aadA2	Streptomycin resistance gene
dfrA12	Trimethoprim resistance gene
intI1	Integron integrase
Unknown	DUF3330 domain-containing protein

Gene Name	Functional annotation (B)
Unknown	Hypothetical protein
repA	Regulatory protein
Unknown	Replication protein C
aph(3')-Ia	Kanamycin resistance gene
Unknown	Hypothetical protein
mobC	Mobilisation protein C
Unknown	Hypothetical protein
mobB	Mobilisation protein B

Gene Name	Functional Annotation (C)
aac(3)-IId	Gentamicin resistance gene
tmrB	Putative tunicamycin resistance protein
Unknown	Hypothetical protein
Unknown	Hypothetical protein

Gene Name (Ordered)	Functional Annotation (D)
qnrS1	Ciprofloxacin resistance gene
Unknown	IS3 family transposase
Unknown	IS3 family transposase
Unknown	IS3 family transposase
Unknown	Tn3 family transposase
Unknown	Hypothetical protein
blaCTX-M-55	Ampicillin and ceftriaxone resistance gene
Unknown	Hypothetical protein
tnpA	IS1380 family transosase
Unknown	Tn3 family transposase
Unknown	Multispecies transposase

Gene Name (Ordered)	Functional Annotation (E)
dgkA	Diacylglycerol kinase
mcr-3.2	Colistin resistance gene
Unknown	NimC/NimA family protein
tnpA	Tn3 family transposase

Figure 4.4 Functional annotation of contigs containing resistance determinants in XDR *S.* Panama isolate

Contiguous sequences containing resistance genes extracted from XDR *S.* Panama. Description of annotations is available in the attached table.

4.4.4 Phylodynamics and evolutionary timescales of *S.* Panama

To infer the evolutionary dynamics and genetic history of the global *S*. Panama population, a Bayesian approach was used (MASCOT template (Müller et al., 2018) implemented through BEAST (Bouckaert et al., 2019) (**Figure 4.5**). Due to the complexity of the evolutionary model (**Data S4.1**) and the limited availability of computational resources, 50/489 *S.* Panama were sub-sampled for the analysis. Because such a small subset was selected to be included in the BEAST analysis, it will be important for future analysis to validate results by taking multiple random subsets of the data and comparing the results of these multiple runs. Alternatively, with increased computational resources, it may be possible to run the analysis on a larger dataset. Importantly, the spatial-temporal reconstruction of phylogeny yielded the same branching pattern as the maximum likelihood analysis, and major clades had greater than 99% Bayesian posterior support (except C3 which contained only 1 isolate following sub-sampling).

The MRCA of all samples in this study date back to the year 1913 (95% HPD = 1815.61-1986.10). This MRCA coincides with the opening of the Panama Canal in 1914 by the USA, raising the possibility of an epidemiological association. I speculate that *S.* Panama may have been circulating in an animal reservoir prior to this time, and the opening of the Panama Canal may have been a window of opportunity for global dissemination in humans. Predictions for MRCA were similar for all clades in this study, specifically 1987 for C1 (95% HPD = 1978.18-1992.75), 1988 for C2 (95% HPD = 1976.18-1996.52), 1988 for C4 (95% HPD = 1945.75-2011.92), 1990 for C5 (95% HPD = 1966.69-2003.99) and 1985 for C6 (95% HPD = 1957.87-2000.99). A limitation of this analysis is that only a subset of randomly selected isolates was used to predict dates for lineage emergence. As previously demonstrated in Chapter 3**,** estimations for MRCA predictions can be improved by using older isolates. If access to historical isolates became available, then it would be beneficial to rerun this analysis to improve the predictions.

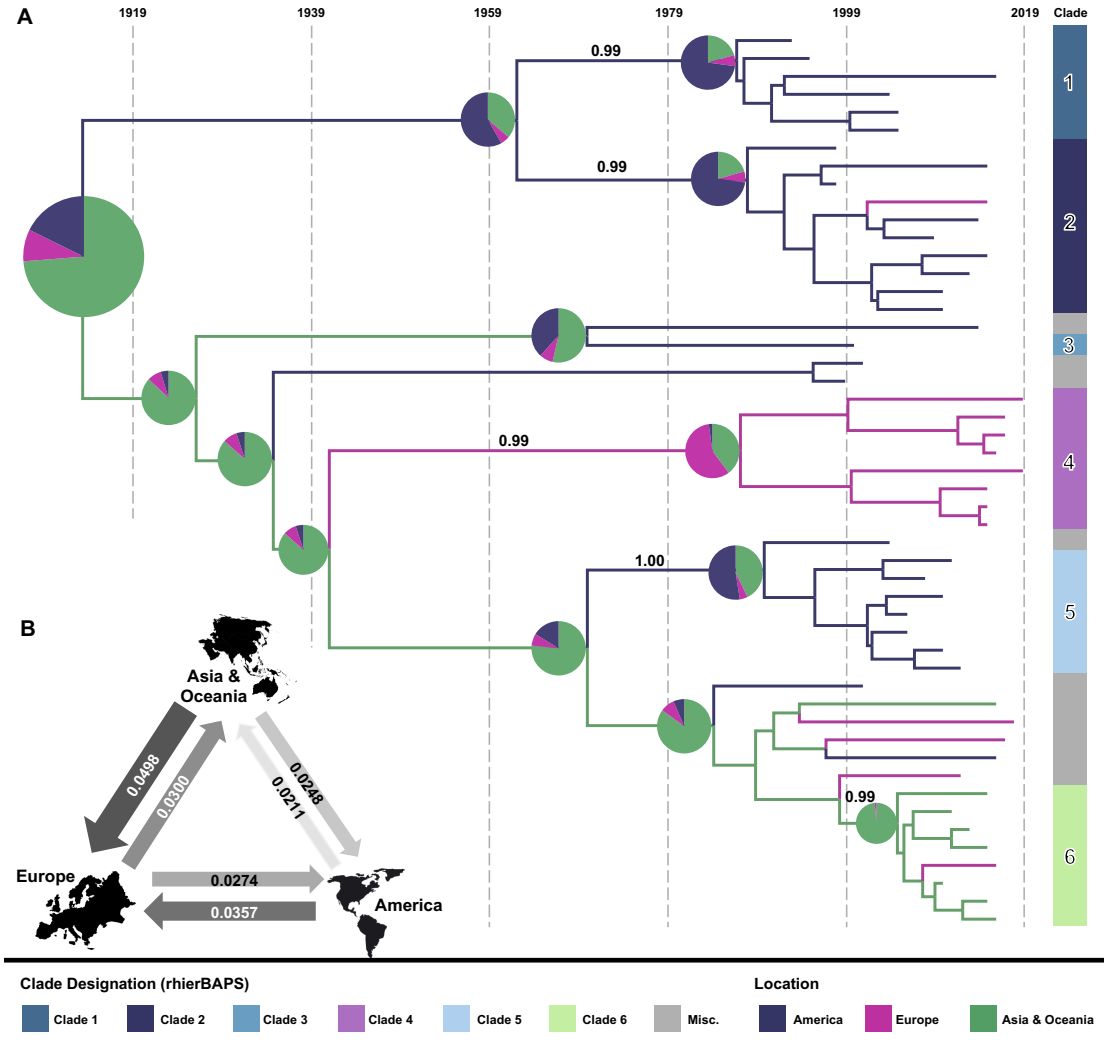


Figure 4.5 Phylogeographic and temporal evolution of *S.* Panama

(A) Chronograph of 50 *S.* Panama strains isolated from America (n=29), Asia/Oceania (n=6) and Europe (n=13). The figure displays a maximum clade credibility tree. The colour of each branch indicates the most likely region of its daughter node. The pie charts indicate the probability of chosen nodes being in any of the possible regions. (B) The mean inferred migration rates for pathogen movement, as indicated by labels and the width and colour of the arrow. The larger the number, the more likely it is that a lineage in the destination location of the arrow originated from the source location of the arrow. Note that the migration rate is displayed in number of migrations per 100 years. Methods available in **Section 4.3.6**.

Analysing bacterial migration has proven useful in the past for establishing the movement of pathogens between different countries (Holt et al., 2012; Okoro et al., 2012). The implementation of a structured coalescent approximation in Bayesian statistics can be used to model the migration process (Barido-Sottani et al., 2017; Müller et al., 2018). To investigate spatial epidemiology, pathogen movement (migrations) was quantified between different geographical regions. The term "migration" is used here to describe the inferred movement of bacterial pathogens from one part of the world to another.

Based on the current dataset, the most recent common ancestor for the *S.* Panama serovar most likely originated from Asia/Oceania (posterior probability Asia/Oceania=0.7376, Europe=0.0825, America=0.1772). This finding is however limited by geographical sampling locations and more specifically an absence of sampling from North America. North American isolates likely play a key role in the evolutionary history of *S.* Panama. Specifically, isolates from the United States (US) contribute 380 of the 912 *S.* Panama genomes available EnteroBase, but these were not included in my analysis. To determine how the US isolates compare with other variants of the *S.* Panama serovar, a phylogenetic tree was generated on EnteroBase. **Figure 4.6** show several US-specific branches of the phylogeny. Future analyses involving the inclusion of samples from North America would be required in order to provide a more complete global picture.

The average migration of *S.* Panama was lowest towards Asia/Oceania, specifically 3.00 migrations per 100 years from Europe (95% HPD = $3.43e^{-5}$ to 9.75) and 2.11 migrations per 100 years from America (95% HPD = $4.34e^{-7}$ to 7.05). The average rate of migration into America was slightly higher, specifically 2.74 migrations per 100 years from Europe (95% HPD = $1.33e^{-5}$ to 10.6) and 2.48 migrations per 100 years from Asia/Oceania (95% HPD = $3.95e^{-5}$ to 8.95). Average migration was greatest into Europe, specifically 4.98 migrations per 100 years from Asia/Oceania (95% HPD = $8.35e^{-4}$ to 12.8) and 3.57 migrations per 100 years from America (95% HPD = $5.11e^{-6}$ to 15.5). The high importation of bacterial strains from Asia into Europe relative to other average migration rates is of particular concern because the majority of Asian isolates carry MDR gene cassettes that confer resistance to multiple classes of antimicrobials.

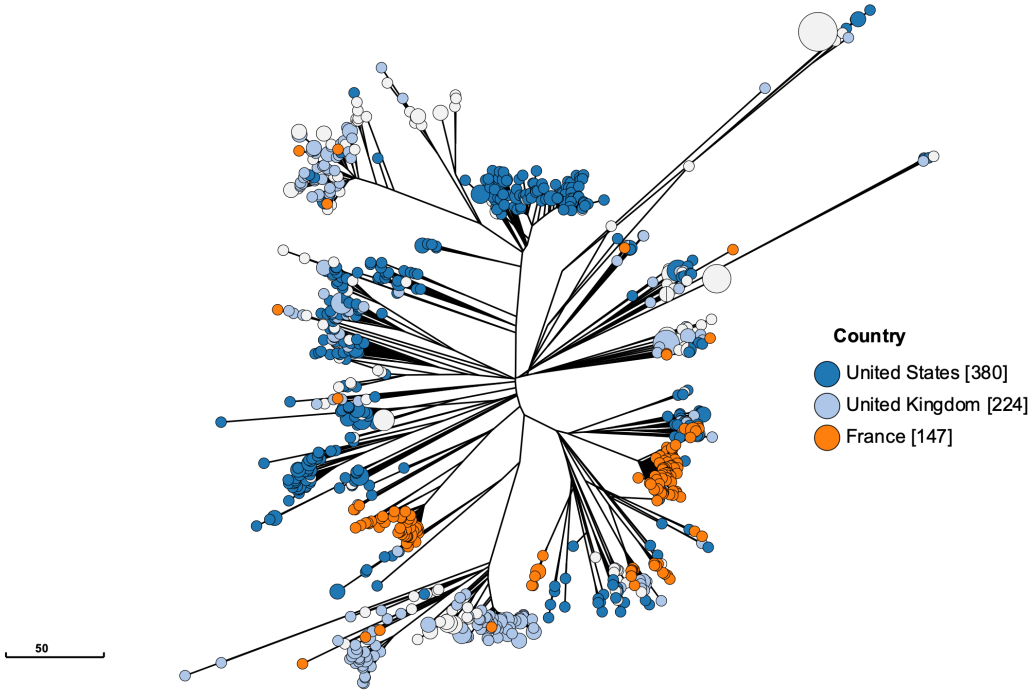


Figure 4.6 cgMLST phylogeny of *S.* Panama genomes available on EnteroBase

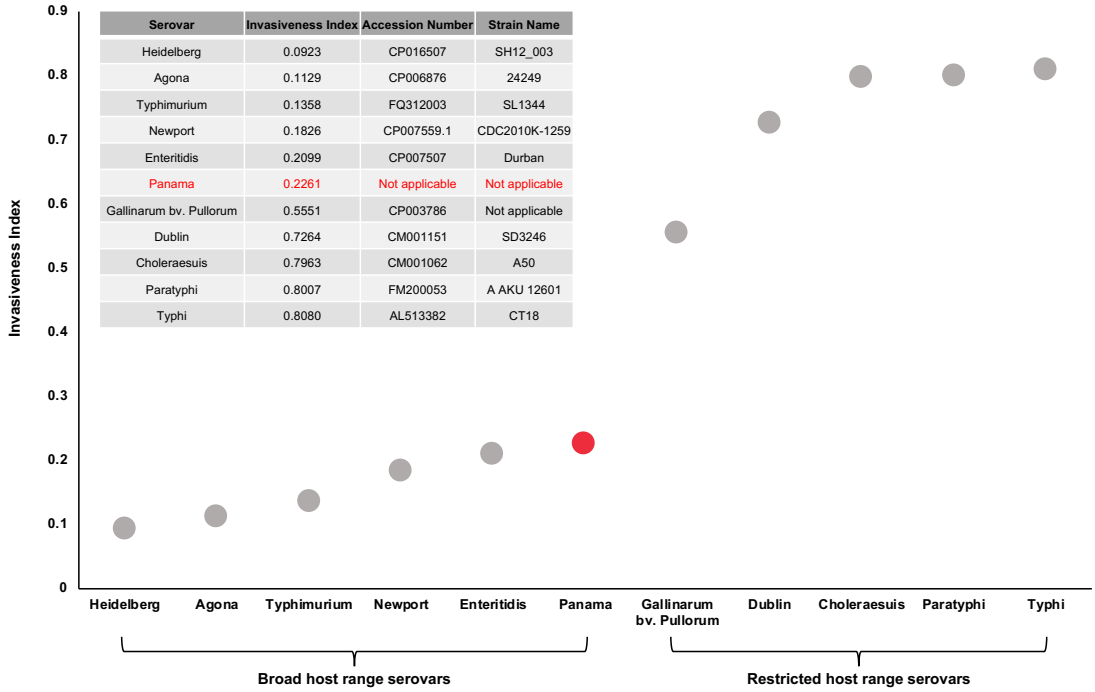
A core genome MLST tree of 912 *S.* Panama (based in cgMLST HC400_369) showing additional United States branches that are not present in the current study. Colours indicate country metadata available on EnteroBase, with grey representing unknown country. The number of isolates available from each country is indicated in square brackets.

4.4.5 The invasiveness index of *S.* Panama

Although *S.* Panama causes extraintestinal disease globally, the genetic basis of the ability of the serovar to cause systemic infection has never been assessed. This capacity, termed invasiveness, reflects the ability of the pathogen to cause systemic spread and extraintestinal infection in humans (Langridge et al., 2012). Invasive infections cause much higher levels of mortality than gastroenteritis; for example, the case-fatality rate of iNTS is 20% (Ao et al., 2015), whereas the case-fatality ratio of *S.* Typhimurium-associated gastroenteritis is 0.6% (Jones et al., 2008). The invasiveness of some *Salmonella* serovars can be assessed experimentally with animal infection models (Tsolis et al., 2011), but no infection model has yet been developed for *S.* Panama. Consequently, a bioinformatics approach was used to determine the relative invasiveness of the various clades of this serovar.

To quantify the extraintestinal potential of *S.* Panama, the invasiveness index was calculated for all 489 isolates in this study. The model previously trained by Wheeler *et al.* (2018) was used, which has already been employed to calculate the invasiveness index of other common *Salmonella* serovars (Van Puyvelde et al., 2019; Wheeler et al., 2018). The model relies upon nucleotide differences within genes identified as predictors for invasiveness (Wheeler et al., 2018). The use of this validated model permitted comparison with the previously described invasiveness indices of other common *Salmonella* serovars (**Figure 4.7**) (Wheeler et al., 2018).

The median invasiveness index of *S.* Panama was determined to be 0.2216 (SD = 0.01952), slightly higher than that described for *S.* Enteritidis, *S.* Newport, *S.* Heidelberg, *S.* Typhimurium and *S.* Agona (Wheeler et al., 2018). In contrast, the invasiveness index of the host-restricted serovars Typhi, Paratyphi, Gallinarum, Pullorum, Dublin and Choleraesuis was significantly higher than S. Panama (Wheeler et al., 2018). SNPs and InDels were identified in each one of the 196 predictor genes for invasiveness, across all samples in this study (when compared to the reference list (Wheeler et al., 2018). Genome degradation is a hallmark feature of host-adapted serovars that cause systemic disease (Nuccio and Bäumler, 2014), raising the possibility that that the *S.* Panama serovar is in the process of evolving towards a restricted host range.



Serovar	Invasiveness Index	Accession Number	Strain Name
Heidelberg	0.0923	CP016507	SH12_003
Agona	0.1129	CP006876	24249
Typhimurium	0.1358	FQ312003	SL1344
Newport	0.1826	CP007559.1	CDC2010K-1259
Enteritidis	0.2099	CP007507	Durban
Panama	0.2261	Not applicable	Not applicable
Gallinarum bv. Pullorum	0.5551	CP003786	Not applicable
Dublin	0.7264	CM001151	SD3246
Choleraesuis	0.7963	CM001062	A50
Paratyphi	0.8007	FM200053	A AKU 12601
Typhi	0.8080	AL513382	CT18

Figure 4.7 Invasiveness index of *S.* Panama compared with common *S. enterica* serovars.

Graph indicates the invasiveness index assigned to *S. enterica* serovars used for approach validation by Wheeler *et al*, (2018) compared with the median invasiveness index of *S.* Panama identified in this study (**Section 4.3.7**).

To determine whether certain *S.* Panama clades had the potential to be more invasive than others, the invasiveness index values of the six *S*. Panama clades was compared, as shown in **Figure 4.8**. The median value of invasiveness index was greatest for C4 (median = 0.2356, SD = 0.01109), followed by C6 (median = 0.2356, SD = 0.01109), C3 (median = 0.2282, SD = 0.01284), C2 (median = 0.2254, SD = 0.01870), C1 (median = 0.2193, SD = 0.01579), and C5 (median = 0.2185, SD = 0.01254). Importantly, the two clades with the highest proportion of AMR isolates had significantly higher invasiveness indices than other clades, highlighting that C4 and C6 could be a future public health issue.

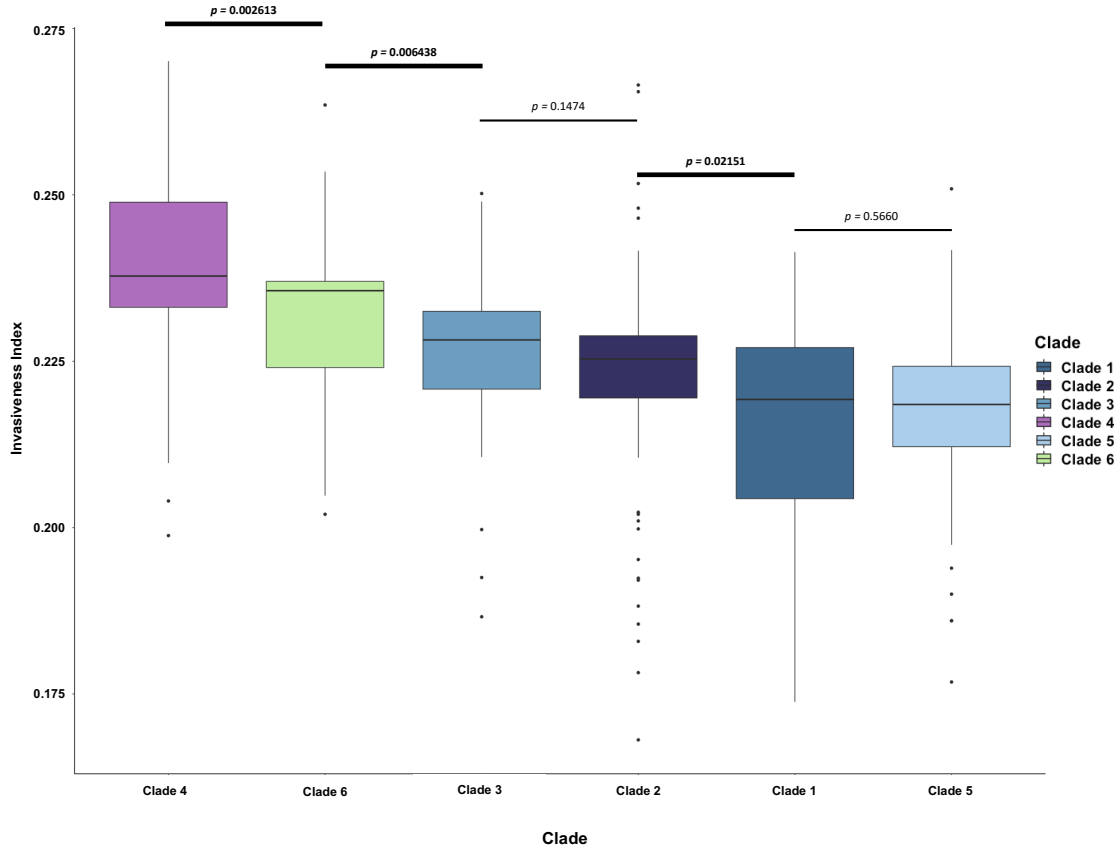


Figure 4.8 Invasiveness index of *S.* Panama clades

Box plot representing the distribution of invasiveness index values for all 489 sequences included in this study summarised by clade assignment (**Section 4.3.7**). To determine whether there was a significant difference between the invasiveness indexes of groups, the Wilcox Mann Whitney test was used and the resultant *p*-values are displayed above groups. Boxplot centre lines represent median values for C1 to C6. Boxplot limits represent upper and lower quartiles, boxplot whiskers represent the 1.5 interquartile range and individual points represent outliers. Boxplot was created using R ggplot2.

4.4.6 Genome wide association study

To understand the genetic basis of the invasiveness of *S.* Panama, a genome-wide association study was conducted. GWAS are increasingly being used to identify the genetic basis of bacterial phenotypic properties (Chen and Shapiro, 2015; Power et al., 2017), and can involve various technical approaches. Because of the advantages and limitations of each method (reviewed elsewhere (Lees, 2017)), a robust GWAS analysis requires the use of a combination of techniques (Lees, 2017; Lees et al., 2018; Lees and Bentley, 2016). Here, two accessory gene-based GWAS approaches were implemented (Brynildsrud et al., 2016; Lees et al., 2018) and one SNP-based GWAS approach (Lees et al., 2018). The results, from each analysis are available in **Table S4.2**. The first accessory gene analysis used a linear fixed effects model to identify 357 significant genes, five of which matched those listed in the top 196 extraintestinal predictor genes used to calculate the invasiveness index, specifically; *rna*, *lyxK*, *mglA*, *pduG, bcsA* and *oppA*. The gene *rna* encodes ribonuclease I and had the lowest *p*-value of all genes in this analysis (lrt *p*-value = $5.09e^{-5}$). The second accessory gene analysis used a pan-GWAS approach to identify 225 significant genes. The hit with the lowest *p*-value corresponded to an unnamed gene predicted to be a lytic transglycosylase (Odds Ratio = 5.74, BH *p*-value = 4.49E-16). Lytic transglycosylases are important in biofilm formation (Monteiro et al., 2011), cell wall metabolism and bacterial pathogenesis (Scheurwater et al., 2008).

To identify the candidate genes most likely to be associated with invasiveness, the findings from the two approaches were cross-referenced. In total, 14 genes were identified, identified by two different GWAS methods that I implemented, and represent good candidates for future invasiveness studies (**Table 4.3**). Four of the 14 genes were located within a prophage region, including a putative site-specific recombinase, a phage tail protein and two hypothetical proteins.

Because individual mutations have previously been found to be responsible for systemic infection (Hammarlöf et al., 2018), the linear fixed effects model (Lees et al., 2018) was also used to search for SNPs associated with the invasive phenotype. No SNPs exceeded the suggestive significance threshold ($p = 1.29e^{-5}$). However, a 46.6kb (1108684-1155312bp) region which represented the highest peak in the Manhattan plot was identified (**Figure 4.9**). The region corresponded the same prophage region that was implicated by the accessory

gene GWAS findings. The prophage was Fels-2-like and shared 72.22% similarity with the prophage RE-2010 (also known as ELPhiS, the Enteritidis lysogenic phage S) (Hanna et al., 2012). RE-2010 is found in many other extraintestinal *Salmonella* serovars, including a global outlier cluster of *Salmonella* Enteritidis associated with foodborne disease (Feasey et al., 2016; Mottawea et al., 2018).

Table 4.3 Genes associated with invasiveness of *S.* Panama, identified by two independent GWAS approaches*

Gene	Annotation	Odds Ratio	BH *p*-value	lrt *p*-value
pin	Putative site-specific recombinase on prophage region	0.032	1.38e-12	2.12e-3
-[a]	Hypothetical protein on prophage region	0.039	2.67e-10	5.04e-3
-[b]	ISNCY family transposase	13.33	1.77e-9	3.23e-2
stf	Phage tail protein on prophage region	0.15	2.73e-8	6.17e-4
-[c]	Hypothetical protein	3.83	4.65e-5	1.24e-2
-[d]	Rpn family recombination-promoting nuclease/putative transposase	6.45	1.45e-4	1.06e-3
-[e]	Rpn family recombination-promoting nuclease/putative transposase	0.26	1.86e-4	1.77e-2
-[f]	Hypothetical protein	5.74	7.56e-4	2.88e-2
-[g]	Hypothetical protein on prophage region	4.12	1.65e-3	2.17e-3
-[h]	tRNA-Lys(ttt)	3.10	1.87e-3	1.44e-2
siiD	Periplasmic adapter subunit of HlyD-family type 1 secretion system involved in adhesion to mammalian cells	0.43	1.00e-2	1.87e-2
-[i]	tRNA-Asp(gtc)	0.28	2.12e-2	4.22e-3
rapA	RNA polymerase	6.29	2.71e-2	1.86e-2
yihQ	alpha glucosidase	0.16	3.06e-2	2.70e-3

* Genes were identified in both Scoary gene presence/absence analysis and Pyseer COG analysis (**Section 4.2.8**)
NB: Nucleotide sequences of these genes are available in **Data S4.2**

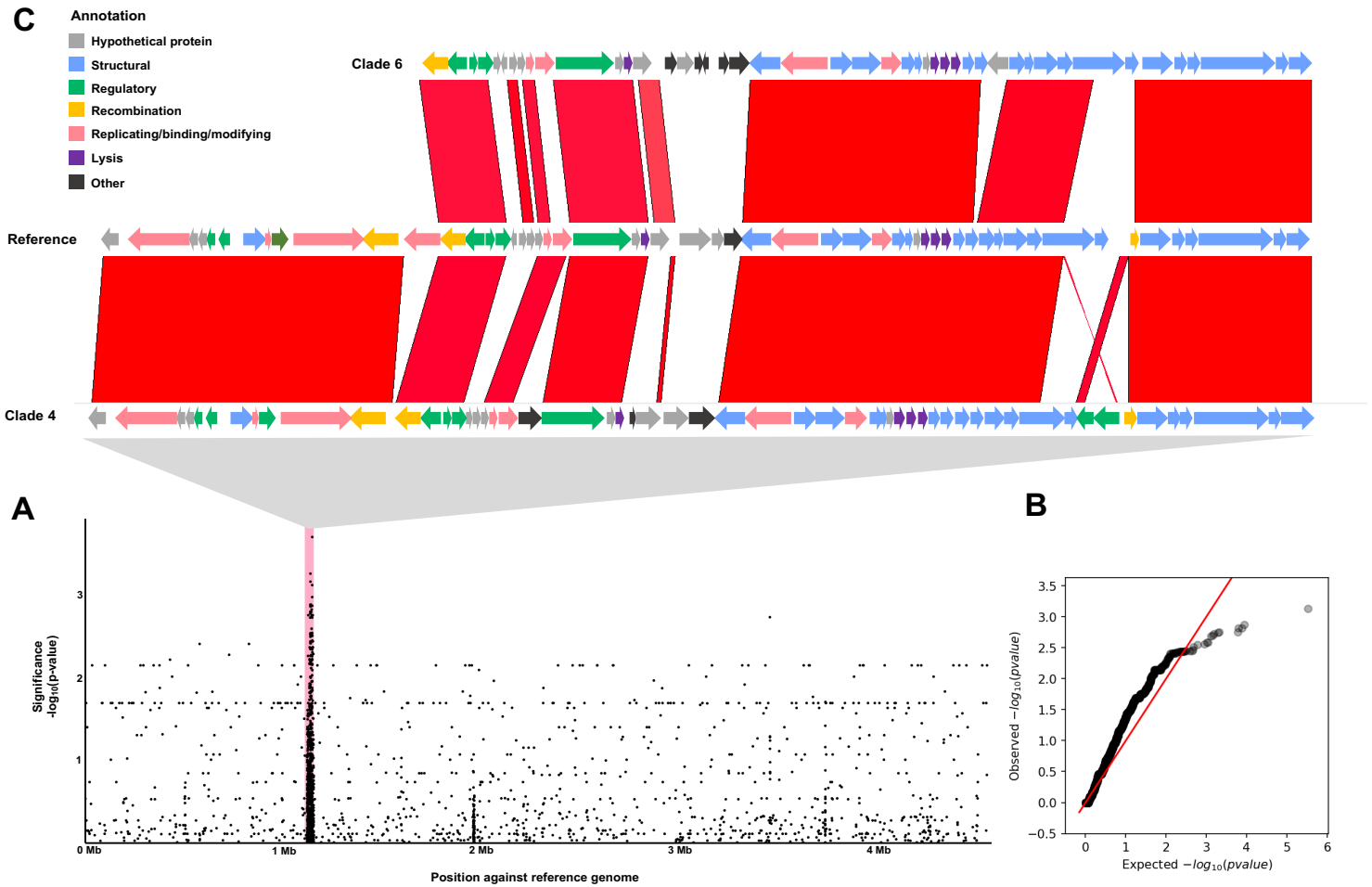


Figure 4.9 GWAS indicating potential contribution of a Fels-2-like-prophage region.
Legend on next page

Figure 4.9 GWAS indicating potential contribution of a Fels-2-like-prophage region.

(A) Manhattan plot showing the –log10 lrt-adjusted *p*-values for SNPs identified in the GWAS (y-axis) plotted against their corresponding position on the reference genome (x-axis) (**Section 4.3.8**). (B) qq plot was created to determine *p*-value inflation and shows expected − log10-transformed *p*-values plotted against observed − log10-transformed *p*-values. (C) Annotation of Fels-2-like prophage region in *S.* Panama reference genome (CP012346), and a representative from C6 and C4. The results of a pairwise comparison between strains is shown as colour blocks, with red indicating presence and white indicating absence (**Section 4.3.9**). Annotation labels are available in **Table S4.3**.

Although invasive-associated SNPs were distributed across the whole Fels-2-like region, the most significant SNPs (with the lowest *p*-values) were associated with two prophage-associated predicted cargo genes. Specifically, SNPs were identified in genes encoding a predicted terminase endonuclease subunit and the RrrD lysozyme. The *rrrD* gene is located in the tail fibre region of the RE-2010 prophage, and over-expression of this gene leads to bacterial cell lysis (Srividhya and Krishnaswamy, 2007). Interestingly, the RrrD protein is required for curli-based biofilm formation (Toba et al., 2011), overlapping with the function of the lytic transglycosylase gene identified in the accessory gene-based GWAS above.

Impaired biofilm formation has been linked to host adaptation and systemic disease in *Salmonella* (MacKenzie et al., 2019; Singletary et al., 2016), and further experimental work would be required to investigate the biofilm-forming properties of *S.* Panama. The most significant SNPs were most commonly found in C4 and C6, indicating lineage effects (**Table S4.2**). Consequently, the conservation of the Fels-2-like prophage region of the *S.* Panama reference strain ATCC 7378 was explored across phylogenetic clades (**Figure 4.10**). I found that the prophage region was more highly conserved in isolates belonging to C4 and C6 than in other clades.

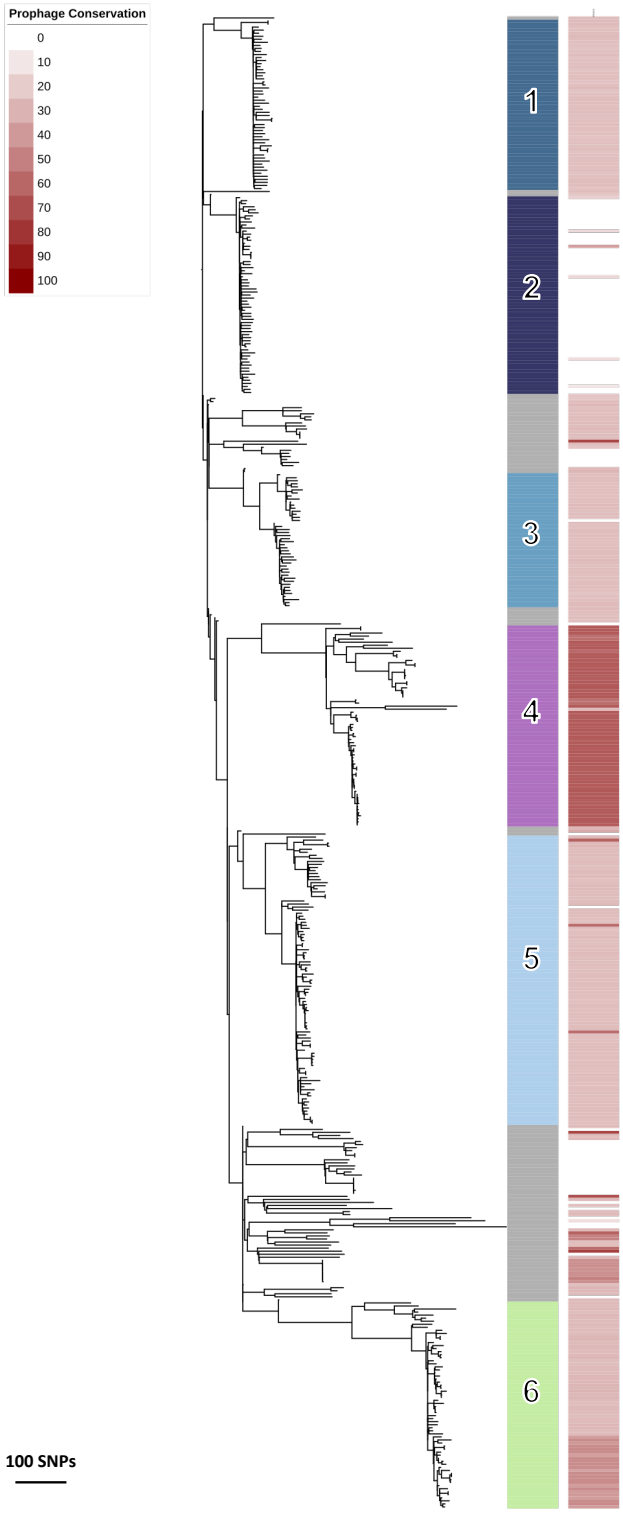


Figure 4.10 Conservation of the Fels-2-like prophage across *S.* Panama clades

Maximum likelihood phylogeny demonstrating the population structure of 489 *S.* Panama (details described in **Figure 4.1**). Left colour strip represents cluster assignment (rHierBAPs), with grey representing a miscellaneous polyphyletic cluster. To visualise the level of conservation of the Fels-2 like prophage in individual isolates, a heat map was used. The right colour strip shows % coverage of the Fels-2-like prophage region in mapped isolates compared to the *S.* Panama reference genome.

4.4.7 Characterisation of prophages carried by *S.* Panama C4 and C6

Prophages are important vehicles for horizontal gene transfer between different *S. enterica* serovars and often carry gene cassettes that contribute to the virulence of host cells (Brussow et al., 2004; Figueroa-Bossi et al., 2001; Ho and Slauch, 2001; Owen et al., 2017; Wagner and Waldor, 2002). To investigate the prophages implicated in invasiveness in *S.* Panama C4 and C6, the respective regions were extracted from a representative sample in each clade and pairwise comparisons were done with well annotated prophages (**Figure 4.11**, **Table S4.3**).

The novel C4 and C6 prophages were predicted to be intact and shared the same attachment site as Fels-2 and SopEΦ (Pelludat et al., 2003), located 3' to the tmRNA-encoding *ssrA* gene. Both of the novel *S.* Panama prophages carried the majority of functional gene clusters required for integration/excision, lysis, and capsid and tail assembly. The C6 prophage was 33.5kb in length, and resembled Fels-2 and RE-2010 in terms of gene content. Functional annotation showed that the C6 prophage also encoded a GNAT family N-acetyltransferase and a putative phage abortive infection system. The C4 prophage was larger (45.7kb) than the C6 prophage and resembled SopEΦ, carrying an additional 12 genes encoding repressor and integrase proteins. The C4 prophage also encoded a phosphoadenosine phosphosulfate reductase family protein and a multidrug ABC transporter.

The discovery of two prophage regions that could be involved in mediating the iNTS phenotype of the two clades with the highest invasiveness index, and the greatest proportion of MDR isolates, clearly warrants further investigation.

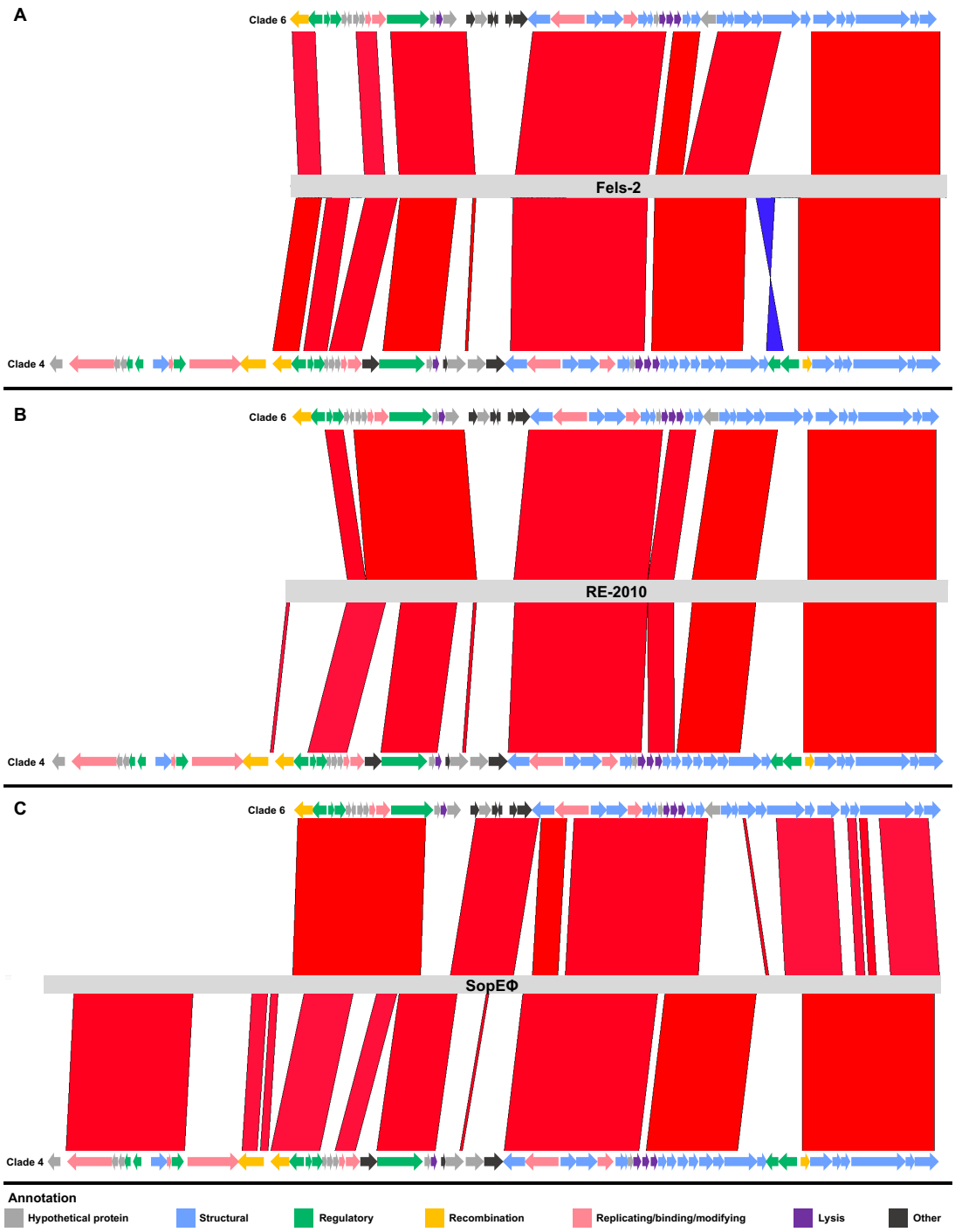


Figure 4.11 Characterisation of Fels-2-like prophage in *S.* Panama C4 and C6

Pairwise comparisons amongst well-characterised prophages and the Fels-2-like prophage in *S.* Panama C4 and C6 (**Section 4.3.9**). Red blocks indicate regions of similarity and white blocks indicate regions absent. Prophage annotations are shown for C4 and C6 (raw data available in **Table S4.3**). Specifically; (A) shows the pairwise comparison amongst C4, Fels-2 (NC_010463.1) and C6; (B) shows the pairwise comparison amongst C4, RE-2010 (HM770079) and C6; (C) shows the pairwise comparison amongst C4, SopEΦ (FQ312003.1 region 2855616 to 2901979) and C6.

4.5 Perspective

This was the first large-scale phylogenetic analysis of the *S.* Panama serovar, and has revealed important information about population structure, AMR, global ecology and invasiveness potential. I have specifically identified six geographically associated *S.* Panama clades which share a most recent common ancestor in Asia/Oceania, and were predicted to have emerged in the early 1900s. By studying the phylogeographical history of *S.* Panama, I have calculated Bayesian migration rates and quantified global pathogen movement. A key finding was the relatively higher rate of immigration of *S.* Panama into Europe, compared with other regions. The importation of Asia/Oceania isolates into Europe could become a threat to public health in the future because the majority of Asian isolates carry MDR gene cassettes that confer resistance to multiple classes of antimicrobials.

I found that the majority of MDR isolates belonged to two phylogenetic clades; C4 (the 'European clade') and C6 ('the Asia/Oceania clade'). Strikingly, C4 and C6 also exhibited the highest invasiveness indices of all clades in this study: the quantification of loss-of-function mutations involving genes which are not necessary in particular environmental conditions (Klemm et al., 2016; McClelland et al., 2004; Nuccio and Bäumler, 2014) suggested an increased extraintestinal potential. Coupled with the elevated invasiveness index, the C4 and C6 clades also carried novel Fels-2-like prophage regions which appeared to carry a signature of invasive disease in my SNP-based GWAS.

It will be important to monitor the incidence of C4 and C6 in ongoing genome-based surveillance. It is possible that the evolutionary trajectory of *S*. Panama will parallel the *Salmonella* serovars Enteritidis and Typhimurium, which contain specific lineages and pathovariants with an increased invasiveness index that cause significant levels of human bloodstream infections (Stanaway et al., 2019a; Van Puyvelde et al., 2019; Wheeler et al., 2018)

Chapter 5

General Discussion

5.1 Brief introduction

The question of how *Salmonella* evolved from simply being a gastrointestinal pathogen to develop the ability to cause systemic infections was first posed decades ago. In the genomic-era of pathogen surveillance and research, WGS-based technologies facilitate the investigation of nontyphoidal *Salmonella* in the context of bloodstream infections. Thus, my thesis aimed to **determine the temporal, geographical and evolutionary trends in nontyphoidal *Salmonella* serovars, in order to understand the genomic signatures acquired on the trajectory towards extraintestinal adaptation.** In this chapter, I will focus on the key conclusions from my research and relate my findings to the important literature.

5.2 Population structure as a predictor for niche-adaptation

In capturing *Salmonella* population-structure at multiple resolutions, I have revealed hitherto unseen insights into the evolution of niche-adaptation in a globally important pathogen. Initially, I found that reptiles were the ideal population of animals for the study of species-level evolution of *S. enterica* because they carry phylogenetically diverse isolates that belong to the majority of *Salmonella* subspecies. Thus, in Chapter 2 the population structure of *S. enterica* was determined using isolates collected from venomous snakes and non-venomous reptiles (**Figure 2.2**). The *S. enterica* subspecies clustered into two clades (A and B) (**Section 2.4.2**), reflecting findings from previous studies based on the alignment of 92 core loci (den Bakker et al., 2011; Falush et al., 2006). An integrated outlook on clade-specific serovar distributions and clinical surveillance data has demonstrated that clade A contains the majority of serovars responsible for salmonellosis in humans, whereas clade B contains several host-adapted serovars (Parsons et al., 2011). Consistent with this epidemiologically suggestive data, I have found clade-specific properties which show the reduced overall metabolic repertoire of clade B (**Section 2.4.5**, **Figure 2.3**), consistent with niche-adaptation that involves a decreased ability to occupy the mammalian gastrointestinal tract. In principle, categorising a serovar as clade B, could be used as a crude predictor for host-adaptation and the ability to cause systemic disease.

S. Panama is an *S. enterica* clade B serovar which fits into this overarching classification, being responsible for extraintestinal infections globally (Pulford et al., 2019). In Chapter 4, I demonstrated that *S.* Panama has a slightly higher invasiveness index than several host-generalist serovars (**Figure 4.7, Figure 5.1),** indicative of a narrowing host range. Global surveillance data suggests that *S.* Panama can survive in various organisms including reptiles, pigs, poultry, cows, goats and wild birds (Cordano and Virgilio, 1996; European Food Safety Authority and European Centre for Disease Prevention and

Control, 2012; Gay et al., 2014; Guyomard-Rabenirina et al., 2019; Matias et al., 2016). It is possible that *S.* Panama is better adapted to occupy these alternative hosts rather than humans, consistent with my finding of that genes involved in survival in the human gut carried loss-of-function mutations (**Section 4.3.5**). These pseudogenisation events are not limited to clade B *Salmonella* and have been seen in other serovars such as *S.* Choleraesuis which is typically associated with swine but causes a severe systemic infection in humans (Chiu et al., 2004). *S.* Choleraesuis is a clade within the Para C sub-lineage of *S. enterica* clade A which also includes *S.* Paratyphi C, Typhisuis and Lomita. Each of these serovars have evolved host-specificity within a specific niche, and are thought to have arisen over the past 4,000 years via a combination of pseudogenes and differential acquisition of two genomic islands (SPI-6 annd SPI-7) according to analysis involving ancient DNA (Zhou et al., 2018). To further elucidate niche-adaptation within *S.* Panama it will be important for future studies to adopt a one-health approach for selecting samples, incorporating samples from various animal reservoirs.

Despite many host-restricted *Salmonella* falling into clade B, some of the most important human-adapted serovars belong to clade A, including *S.* Typhi and *S.* Paratyphi C (den Bakker et al., 2011). Thus, classifying a serovar as clade A does not necessarily reflect host-range and disease outcome. It can also be the case that diversifying selection may occur within serovars that typically occupy a broad range of hosts, leading to independent lineages that have evolved to exploit a specific niche. For example, in Chapter 3 I have uncovered a series of events that have helped to shape the emergence of three invasive *S.* Typhimurium ST313 lineages in Africa, each at a different stage of niche adaptation.

My data revealed the process of sequential genome decay and accessory genome acquisition, highlighting the evolutionary events that were crucial for the development of each lineage (**Figure 3.12**). For example, the pseudogenisation of *macB, sseI* and *lpxO* appear to be critical to the emergence of ST313 L2 (**Section 3.4.6**). The stepwise pseudogenisation of genes reported here is consistent with the host adaptation of *S*. Typhimurium ST313 lineages over evolutionary history and supports mounting evidence for a human ST313 reservoir (Post et al., 2019). My findings in Chapter 3 relied upon lineage-specific evolutionary insights and phylodynamic investigations, demonstrating that it is no longer enough to understand niche-adaptation based purely on serovar designations.

5.3 Using genome degradation to predict invasiveness

I found a common theme of pseudogenisation in iNTS serovars throughout this thesis, supporting previous findings that genome degradation can be used as a marker for invasive disease (Langridge et al., 2015). The invasiveness index is a good tool for quantifying genome degradation and predicting extraintestinal potential (Wheeler et al., 2018). In Chapter 3, I used the invasiveness index to demonstrate that ST313 L3 has an elevated extraintestinal potential compared with that of ST19, ST313 L1, ST313 L2 and UK-ST313 (**Section 3.4.6**, **Figure 3.13**). Consistent with these findings, ST313 L3 was also found to contain additional degradation in genes required for colonisation of the gastrointestinal tract by *S.* Typhimurium (**Section 3.4.7**).

The invasiveness index was also used in Chapter 4, to characterise *S.* Panama. All of the *S.* Panama clades had a higher invasiveness index than the ST313 lineages, highlighting the increased extraintestinal potential of the *S.* Panama serovar compared with the *S.* Typhimurium serovar (**Figure 5.1**). Amongst *S.* Panama, C4 and C6 had an elevated index (**Section 4.3.5, Figure 4.8**). It is possible that the evolutionary trajectory of *S*. Panama will parallel *S.* Typhimurium and *S.* Enteritidis, containing specific lineages and pathovariants with an increased invasiveness index and causing significant levels of human bloodstream infections (Stanaway et al., 2019a; Van Puyvelde et al., 2019; Wheeler et al., 2018). Thus, it will be important to monitor the incidence of C4 and C6 in ongoing genome-based surveillance.

Clearly, the use of genome-based genetic signatures does not prove with absolute certainty that particular iNTS pathovariants have adapted to an extraintestinal lifestyle. In the past, infection models have been important for understanding the virulence and systemic spread of *Salmonella* pathovariants that cause gastroenteritis (Tsolis et al., 2011). However, novel or uncharacterised *Salmonella* pathovariants generally lack animal models that accurately recapitulates human disease. Furthermore, the close relationship between systemic disease and host-adaptation has prompted concerns about the use of animals to assess the levels of invasiveness in terms of humans. A similar issue arose for the invasive pathogen *Salmonella* Typhi, leading researchers to develop a human infection model that proved that the typhoid toxin is not required for clinical infection (Gibani et al., 2019).

No human challenge model has yet been developed for *S.* Panama or *S.* Typhimurium ST313, and previous studies have demonstrated that cellular and animal infection models fail to discriminate

between levels of invasiveness of ST313 L1 and ST313 L2 (Okoro et al., 2015). This problem was exemplified in a recent study which defined a new ST313 sub-lineage identified in the DRC. The study by Van Puyvelde and colleagues (Van Puyvelde et al., 2019) showed that the murine whole-animal infection model and the human macrophage cellular infection model did not find robust differences in terms of invasive disease between the different ST313 lineages and sub-lineages. Extensive animal experiments in chickens, chicken embryos and *Galleria* larvae (Lacharme-Lora et al., 2019) have revealed the limitations of discriminating between ST19 and ST313 in all three systems.

Although the invasiveness index cannot be experimentally validated at the moment, it is clear that the *Salmonella* isolates that vary in terms of invasiveness index do produce distinct clinical symptoms in a human population (Langridge et al., 2009a; Tack et al., 2020). The routine monitoring of genome degradation could be beneficial in a public health setting, to identify host-adapted *Salmonella* variants likely to be responsible for future outbreaks of invasive disease at an early stage. This technique would prove particularly useful for emerging pathovariants, for which no animal model yet exists, such as ST313 (Chapter 3) and *S.* Panama (Chapter 4). A limitation of this approach would be the potential for false negatives, which may mean that invasive serovars which do not exhibit genome degradation are underestimated in terms of clinical disease severity. For example, an internationally distributed monophasic *S.* Typhimurium ST34 variant (*S*. I:4,[5],12:i:−) has recently reacquired a secondary flagellin gene and become associated with invasive disease in HIV-infected individuals. Unlike ST313 in sub-Saharan Africa, this Vietnamese ST34 variant does not exhibit extensive genome degradation (Mather et al., 2018). The development and validation of experimental approaches for the robust measurement of the invasive potential of iNTS isolates will be an important focus for future research.

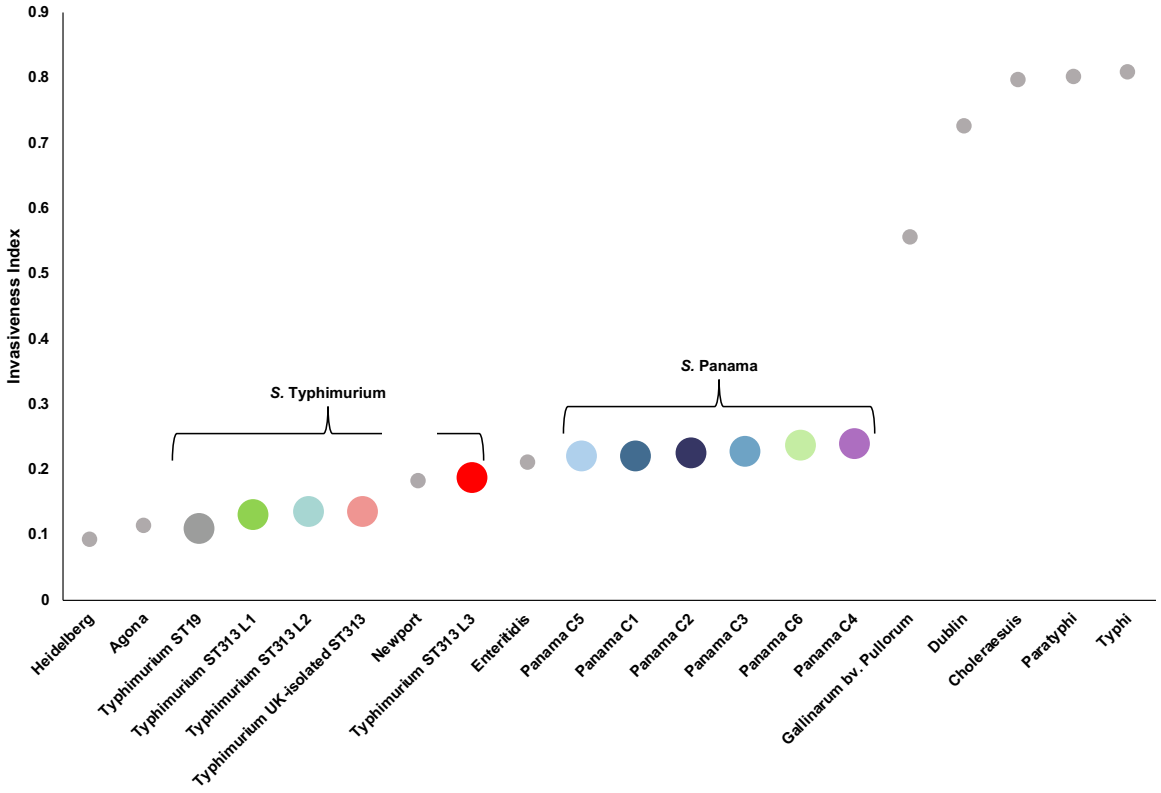


Figure 5.1 Invasiveness index of *S*. Typhimurium and *S*. Panama compared with common *S. enterica* serovars.

Graph indicates the invasiveness index assigned to *S. enterica* serovars used for approach validation by Wheeler *et al*. (2018) (smaller circles) compared with the median invasiveness index of *S.* Typhimurium lineages (**Section 3.3.9**) and *S.* Panama clades (**Section 4.3.7**) identified in this study (larger circles). Colours are used to distinguish between different lineages and clades. **Figure adapted from Chapter 4, Figure 4.7.**

5.4 AMR status likely reflects the selection pressures acting upon *Salmonella* pathovariants

The evolution of virulence to *Salmonella* lineages involves more than just genome degradation. Another factor that has a profound impact upon the emergence of successful iNTS lineages is AMR, a property that can be horizontally transferred between bacteria. Plasmid-mediated conjugation of AMR genes has recently been demonstrated experimentally following exposure of *S.* Typhimurium to antibiotics in the mammalian gut (Bakkeren et al., 2019).

Differing AMR profiles provide an indication of alternative origins of bacteria and/or variation in selective evolutionary pressures within and between bacterial niches (Liao et al., 2020; Mellor et al., 2019). Throughout this thesis, I have presented evidence of AMR reflecting environmental selection pressures. For example, I found very low levels of antibiotic resistance in the *Salmonella* isolated from venomous snakes housed in the LSTM venom unit (**Section 2.4.3**, **Table 2.4**), consistent with a lack of selection pressure associated with the limited use of antibiotics in reptiles.

I developed this theme by using a genomic approach to confirm that clades of *S.* Panama from French territories of America such as Guadeloupe do not usually carry AMR genes (**Section 4.3.3**, **Figure 4.1**). Wild reptiles have previously been described as a key reservoir for *S.* Panama infections in these regions of America (Guyomard-Rabenirina et al., 2019) and as I observed in venomous snakes, the pan-susceptibility to antibiotics likely reflects a low selection pressure in the wild reptilian reservoirs in the French Caribbean.

In contrast, the majority of MDR *S.* Panama in this study originated in Asia/Oceania or Europe (**Section 4.3.3**, **Figure 4.1**). Microorganisms associated with Asia tend to have higher rates of resistance (Nguyen et al., 2013; Nhung et al., 2016). Heavy use of antimicrobials has driven the circulation of numerous MDR *Salmonella* lineages in Asia, including a new variant of MDR *S.* Typhimurium ST34 which has been associated with iNTS in Vietnamese HIV-infected patients (Mather et al., 2018) and an XDR *S.* Typhi clone encoding resistance to fluoroquinolones and third-generation cephalosporins (Klemm et al., 2018). Thus, *S.* Panama represents yet another example of the widespread AMR crisis in Asia.

In **Section 4.3.4**, the inter-regional migration patterns demonstrated pathogen movement from Asia/Oceania into Europe, which may have contributed to AMR in the European *S.* Panama clade (C4) (**Figure 4.5**). However, this European clade (C4) may also have arisen following an outbreak that involved the continental spread of *S.* Panama across Europe, associated with the food production industry in the 1970's and 1980's (Cherubin, 1981; Le Minor and Le Minor, 1981; Lee, 1974; Wilkins and Roberts, 1988). It likely that the European lineage of *S*. Panama (C4) has been exposed to a high antibiotic selective pressure in humans or food animals and consequently became resistant to multiple antibiotic agents via acquisition of multiple mobile genetic elements.

Because AMR *Salmonella* infections are associated with a higher probability of hospitalisation and treatment failure (Krueger et al., 2014), there has been a general bias towards sequencing MDR isolates in human surveillance programmes. However, such a bias may obscure the evolutionary drivers of AMR that can only be revealed by studying the full range of resistance profiles, including susceptible isolates. In Chapter 3, sub-sampling of a comprehensive archive of pathogens allowed an extensive range of AMR profiles in *S.* Typhimurium ST313 responsible for bloodstream infections to be captured (**Section 3.3.4, Figure 3.7**). Across the sampling period changes in antimicrobial usage policies occurred at the local level (Queen Elizabeth Central Hospital) in Malawi, including the phased removal of chloramphenicol from clinical practice. From approximately 2002 onwards, chloramphenicol was replaced by the oral fluoroquinolone, ciprofloxacin for treatment of culture-confirmed iNTS disease. Additionally, beginning in 2005, the Department of Medicine Malawi began to use the third-generation cephalosporin, ceftriaxone for the empirical management of suspected sepsis (Feasey et al., 2015). More recently, chloramphenicol has largely been removed from hospital and private pharmacies in Malawi suggesting that it is infrequently used to treat any infection (M. Gordon, personal communication). Across the study sampling period a reduction in the proportion of chloramphenicol-resistant isolates of ST313 L2 was observed (**Figure 3.7**), raising the possibility that a change in policy on empirical chloramphenicol use played a role in shaping the epidemiology of ST313 *S*. Typhimurium in Malawi.

Combining the available details on antimicrobial usage with fluctuating AMR profiles allows speculation on trends in circulating *Salmonella* lineages, particularly the switch from MDR ST313 L2 to the pan-susceptible ST313 L3 in (**Section 3.4.2**). Interestingly, UK-isolated ST313 also tend to be susceptible to antimicrobial agents (Ashton et al., 2017), raising the possibility that ST313 L3 was introduced to Malawi from the UK via an international transmission event. Furthermore, I hypothesise

that the phased removal of chloramphenicol from clinical practice following Malawian policy changes has reduced selection pressure and opened a window of opportunity for the emergence of fully susceptible ST313 L3 in Blantyre. A comprehensive epidemiological study on antimicrobial usage would be required to properly investigate this hypothesis.

Combining details of resistance determinants with population structure demonstrates the significant advantages of using WGS to study AMR. Previously, such studies have permitted the identification of distinct *S.* Typhimurium (Definitive Type) DT104 populations found in co-located animals and humans in an industrialised country (Mather et al., 2013) and discriminated between *S.* Enteritidis lineages associated with gastrointestinal disease in high income settings compared with invasive disease in low-income settings (Feasey et al., 2016). Studies in other bacterial pathogens have demonstrated how horizontal transfer of an AMR plasmid amongst *Shigella* enhanced existing epidemics and facilitated new outbreaks in the UK (Baker et al., 2018b). Taken together, my findings on AMR have captured a snapshot of population-level epidemiological consequences of AMR in the *Salmonella* population, highlighting how antibiotic selective pressures cause fundamental shifts in the circulation of lineages.

5.5 Parallel prophage repertoires of invasive pathovariants

As important vehicles for horizontal gene transfer, prophages also play a diverse and complex role in the evolution of bacterial pathogens in response to external selection pressures and often carry gene cassettes that contribute to the virulence of host bacterial cells (Brussow et al., 2004; Fang et al., 2017; Figueroa-Bossi et al., 2001; Ho and Slauch, 2001; Owen et al., 2017; Wagner and Waldor, 2002). For example, ST313 L1 and L2 carry a unique prophage repertoire that includes BTP1 and BTP5. In Chapter 3, I have demonstrated that both of these prophages were absent from ST313 L3 and UK-isolated ST313 (**Section 3.4.3**, **Figure 3.4**), suggesting that the two prophages were independently acquired by L1 and L2, an important example of convergent evolution.

BTP1 has recently been found to carry the BstA family of prophage defence proteins which defend bacterial cells against exogenous attack by a variety of lytic phage, without sacrificing lytic autonomy, thus conferring a beneficial trait (Owen et al., 2020). ST313 L3 does not carry BTP1, but instead carries a similar P22-like prophage which differs in the phage immunity region, carrying the *sieB* superinfection exclusion phage immunity protein (**Section 3.4.3**) (Susskind et al., 1974). It is possible that the carriage of a gene that mediates abortive phage infection confers a selective advantage to ST313 L3, just as BstA improves the fitness of ST313 L2 during phage attack (Owen et al., 2020).

An unexpected finding from my study was the identification of Fels-2-like prophages, in *S.* Panama C4 and C6 and *S.* Typhimurium ST313 L3 which were genetically similar to RE-2010 (**Section 3.4.3**, **Section 4.3.7**, **Figure 4.11**). The RE-2010 prophage has also been identified in *S*. Enteritidis causing iNTS disease globally (Feasey et al., 2016). Previously, functional insights into Fels-2-like prophages have revealed important virulence genes. For example the *sopE* gene found in the tail fibre region of the prophage sopEΦ encodes a protein involved in modulating host cell signalling and causes dramatic increases of bacterial cell invasion into host cells (Pelludat et al., 2003). Although the *sopE* gene was not present in any of the Fel-2-like prophages in this study, other genes were highlighted as potentially associated with invasiveness by my GWAS (**Section 4.3.6**, **Figure 4.9**). The discovery of closely related prophages in the three lineages/clades predicted to have the highest invasiveness potentials throughout this thesis clearly warrants further investigations at the functional level, particularly focussing on genes highlighted in my GWAS.

5.6 Towards an integrated approach for pathogen surveillance

My thesis has demonstrated the benefit of using genomic epidemiology to understand nontyphoidal *Salmonella* serovars and their association with invasive disease, showcasing the revolutionary impact of WGS-based approaches. Many of the samples involved in this study were sourced from low-middle income settings, which have modest accessibility to routine WGS for pathogen surveillance. Thus, my study has been important in improving our understanding of the genomic epidemiology of iNTS in these regions. For the future, it is important to strive to improve the accessibility of genomic research for lower income settings by continuing studies such as this and boosting capacity.

As genomics becomes a cornerstone of contemporary pathogen surveillance and prevention, it will be important to maintain awareness of the limitations of computational approaches. I look forward to an integrated approach to public health and scientific research that will enhance our understanding of bacterial pathogens in the future. Harnessing the combined power of genome-level discrimination, with epidemiological rigour and clinical observation with experimental biology will, ultimately, benefit human health.

Bibliography

Achtman, M., Wain, J., Weill, F.-X., Nair, S., Zhou, Z., Sangal, V., Krauland, M.G., Hale, J.L., Harbottle, H., Uesbeck, A., Dougan, G., Harrison, L.H., Brisse, S., The S enterica MLST study group, 2012. Multilocus Sequence Typing as a Replacement for Serotyping in Salmonella enterica. PLOS Pathog. 8, e1002776. https://doi.org/https://doi.org/10.1371/journal.ppat.1002776

Alcock, B.P., Raphenya, A.R., Lau, T.T.Y., Tsang, K.K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L. V, Cheng, A.A., Liu, S., Min, S.Y., Miroshnichenko, A., Tran, H.-K., Werfalli, R.E., Nasir, J.A., Oloni, M., Speicher, D.J., Florescu, A., Singh, B., Faltyn, M., Hernandez-Koutoucheva, A., Sharma, A.N., Bordeleau, E., Pawlowski, A.C., Zubyk, H.L., Dooley, D., Griffiths, E., Maguire, F., Winsor, G.L., Beiko, R.G., Brinkman, F.S.L., Hsiao, W.W.L., Domselaar, G. V, McArthur, A.G., 2019. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 48, D517–D525. https://doi.org/10.1093/nar/gkz935

Alikhan, N.-F., Zhou, Z., Sergeant, M.J., Achtman, M., 2018. A genomic overview of the population structure of Salmonella. PLOS Genet. 14, e1007261. https://doi.org/10.1371/journal.pgen.1007261

Almeida, F., Seribelli, A.A., da Silva, P., Medeiros, M.I.C., dos Prazeres Rodrigues, D., Moreira, C.G., Allard, M.W., Falcão, J.P., 2017. Multilocus sequence typing of Salmonella Typhimurium reveals the presence of the highly invasive ST313 in Brazil. Infect. Genet. Evol. 51, 41–44. https://doi.org/10.1016/j.meegid.2017.03.009

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anderson, S., 1981. Shotgun DNA sequencing using cloned DNase I-generated fragments. Nucleic Acids Res. 9, 3015–3027. https://doi.org/10.1093/nar/9.13.3015

Andrews, J.M., Howe, R.A., 2011. BSAC standardized disc susceptibility testing method (version 10). J. Antimicrob. Chemother. 66, 2726–2757. https://doi.org/10.1093/jac/dkr359

Anon, 2013. Eight Multistate Outbreaks of human Salmonella infections linked to small turtles- United States 2013. [WWW Document]. Centers Dis. Control Prev. URL www.cdc.gov/salmonella/small-turtles-03-12/index.html (accessed 6.9.18).

Antunes, P., Mourão, J., Pestana, N., Peixe, L., 2011. Leakage of emerging clinically relevant multidrug-resistant Salmonella clones from pig farms. J. Antimicrob. Chemother. 66, 2028–2032. https://doi.org/10.1093/jac/dkr228

Ao, T.T., Feasey, N.A., Gordon, M.A., Keddy, K.H., Angulo, F.J., Crump, J.A., 2015. Global Burden of Invasive Nontyphoidal Salmonella Disease, 20101. Emerg. Infect. Dis. 21, 941–949. https://doi.org/10.3201/eid2106.140999

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44, W16-21. https://doi.org/10.1093/nar/gkw387

Arredondo-Alonso, S., Willems, R.J., van Schaik, W., Schürch, A.C., 2017. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. Microb. Genomics 3, e000128–e000128. https://doi.org/10.1099/mgen.0.000128

Ashton, P., Nair, S., Peters, T., Tewolde, R., Day, M., Doumith, M., Green, J., Jenkins, C., Underwood, A., Arnold, C., de Pinna, E., Dallman, T., Grant, K., 2015. Revolutionising Public Health Reference Microbiology using Whole Genome Sequencing: Salmonella as an exemplar. bioRxiv. https://doi.org/https://doi.org/10.1101/033225

Ashton, P.M., Nair, S., Peters, T.M., Bale, J.A., Powell, D.G., Painset, A., Tewolde, R., Schaefer, U., Jenkins, C., Dallman, T.J., de Pinna, E.M., Grant, K.A., 2016. Identification of Salmonella for public health surveillance using whole genome sequencing. PeerJ 4, e1752. https://doi.org/10.7717/peerj.1752

Ashton, P.M., Owen, S. V, Kaindama, L., Rowe, W.P.M., Lane, C.R., Larkin, L., Nair, S., Jenkins, C., de Pinna,

E.M., Feasey, N.A., Hinton, J.C.D., Dallman, T.J., 2017. Public health surveillance in the UK revolutionises our understanding of the invasive Salmonella Typhimurium epidemic in Africa. Genome Med. 9, 92. https://doi.org/10.1186/s13073-017-0480-7

Aulicino, A., Rue-Albrecht, K.C., Preciado-Llanes, L., Napolitani, G., Ashley, N., Cribbs, A., Koth, J., Lagerholm, B.C., Ambrose, T., Gordon, M.A., Sims, D., Simmons, A., 2018. Invasive Salmonella exploits divergent immune evasion strategies in infected and bystander dendritic cell subsets. Nat. Commun. 9, 4883. https://doi.org/10.1038/s41467-018-07329-0

Avril, J.L., Dabernat, H.J., Gerbaud, G.R., Horodniceanu, T., Lambert-Zechovsky, N., Le Minor, S., Mendez, B., Chabbert, Y.A., 1977. [R plasmids incompatibility groups in epidemic Salmonella (author's transl)]. Ann. Microbiol. (Paris). 128, 165–175.

Baggesen, D., Sørensen, G., Moller Nielsen, E., Wegener, H., 2010. Phage typing of Salmonella Typhimurium: Is it still a useful tool for surveillance and outbreak investigation? Euro Surveill. Eur. Commun. Dis. Bull. 15, 19471.

Baker, K.S., Dallman, T.J., Field, N., Childs, T., Mitchell, H., Day, M., Weill, F.-X., Lefèvre, S., Tourdjman, M., Hughes, G., Jenkins, C., Thomson, N., 2018a. Genomic epidemiology of Shigella in the United Kingdom shows transmission of pathogen sublineages and determinants of antimicrobial resistance. Sci. Rep. 8, 7389. https://doi.org/10.1038/s41598-018-25764-3

Baker, K.S., Dallman, T.J., Field, N., Childs, T., Mitchell, H., Day, M., Weill, F.-X., Lefèvre, S., Tourdjman, M., Hughes, G., Jenkins, C., Thomson, N., 2018b. Horizontal antimicrobial resistance transfer drives epidemics of multiple Shigella species. Nat. Commun. 9, 1462. https://doi.org/10.1038/s41467-018-03949-8

Bakkeren, E., Huisman, J.S., Fattinger, S.A., Hausmann, A., Furter, M., Egli, A., Slack, E., Sellin, M.E., Bonhoeffer, S., Regoes, R.R., Diard, M., Hardt, W.-D., 2019. Salmonella persisters promote the spread of antibiotic resistance plasmids in the gut. Nature 573, 276–280. https://doi.org/10.1038/s41586-019-1521-8

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V, Sirotkin, A. V, Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021

Barczi, J., Foldes, G., 1990. [Septic arthritis associated with childhood salmonellosis]. Orv. Hetil. 131, 979–980.

Barido-Sottani, J., Bošková, V., Plessis, L. Du, Kühnert, D., Magnus, C., Mitov, V., Müller, N.F., PečErska, J., Rasmussen, D.A., Zhang, C., Drummond, A.J., Heath, T.A., Pybus, O.G., Vaughan, T.G., Stadler, T., 2017. Taming the BEAST—A Community Teaching Material Resource for BEAST 2. Syst. Biol. 67, 170–174. https://doi.org/10.1093/sysbio/syx060

Battistuzzi, F.U., Feijao, A., Hedges, S.B., 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. BMC Evol. Biol. 4, 44. https://doi.org/10.1186/1471-2148-4-44

Bäumler, A., Fang, F.C., 2013. Host specificity of bacterial pathogens. Cold Spring Harb. Perspect. Med. 3, a010041. https://doi.org/10.1101/cshperspect.a010041

Bäumler, A.J., Tsolis, R.M., Ficht, T.A., Adams, L.G., 1998. Evolution of Host Adaptation in Salmonella enterica. Infect. Immun. 66, 4579–4587. https://doi.org/10.1128/IAI.66.10.4579-4587.1998

Beltran, P., Musser, J.M., Helmuth, R., Farmer, J.J. 3rd, Frerichs, W.M., Wachsmuth, I.K., Ferris, K., McWhorter, A.C., Wells, J.G., Cravioto, A., 1988. Toward a population genetic analysis of Salmonella: genetic diversity and relationships among strains of serotypes S. choleraesuis, S. derby, S. dublin, S. enteritidis, S. heidelberg, S. infantis, S. newport, and S. typhimurium. Proc. Natl. Acad. Sci. U. S. A. 85, 7753–7757.

Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B 57, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bennett, S., 2004. Solexa Ltd. Pharmacogenomics 5, 433–438. https://doi.org/10.1517/14622416.5.4.433

Best, E.L., Lindstedt, B.-A., Cook, A., Clifton Hadley, F.A., Threlfall, E.J., Liebana, E., 2007. Multiple-locus variable-number tandem repeat analysis of Salmonella enterica subsp. enterica serovar Typhimurium:

comparison of isolates from pigs, poultry and cases of human gastroenteritis. J. Appl. Microbiol. 103, 565–572. https://doi.org/10.1111/j.1365-2672.2007.03278.x

Betancor, L., Pereira, M., Martinez, A., Giossa, G., Fookes, M., Flores, K., Barrios, P., Repiso, V., Vignoli, R., Cordeiro, N., Algorta, G., Thomson, N., Maskell, D., Schelotto, F., Chabalgoity, J.A., 2010. Prevalence of Salmonella enterica in Poultry and Eggs in Uruguay during an Epidemic Due to Salmonella enterica Serovar Enteritidis. J. Clin. Microbiol. 48, 2413–2423. https://doi.org/10.1128/JCM.02137-09

Betancor, L., Yim, L., Martínez, A., Fookes, M., Sasias, S., Schelotto, F., Thomson, N., Maskell, D., Chabalgoity, J.A., 2012. Genomic Comparison of the Closely Related Salmonella enterica Serovars Enteritidis and Dublin. Open Microbiol. J. 6, 5–13. https://doi.org/10.2174/1874285801206010005

Blattner, F.R., Plunkett, G. 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997. The complete genome sequence of Escherichia coli K-12. Science 277, 1453–1462.

Bogomolnaya, L.M., Andrews, K.D., Talamantes, M., Maple, A., Ragoza, Y., Vazquez-Torres, A., Andrews-Polymenis, H., 2013. The ABC-Type Efflux Pump MacAB Protects Salmonella enterica serovar Typhimurium from Oxidative Stress. MBio 4, e00630-13. https://doi.org/10.1128/mBio.00630-13

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Borderon, J.C., Prieur, D., Huguet, B., 1981. [Cefotaxime CSF levels in children with purulent meningitis (author's transl)]. Nouv. Presse Med. 10, 580–584.

Bouanchaud, D.H., Chabbert, Y.A., 1969. Stable coexistence of three resistance factors (fi-) in Salmonella panama and Escherichia coli K12. J. Gen. Microbiol. 58, 107–113. https://doi.org/10.1099/00221287-58-1-107

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T., Drummond, A.J., 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLOS Comput. Biol. 15, e1006650.

Bouckaert, R.R., 2010. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics 26, 1372–1373. https://doi.org/10.1093/bioinformatics/btq110

Bouckaert, R.R., Drummond, A.J., 2017. bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evol. Biol. 17, 42. https://doi.org/10.1186/s12862-017-0890-6

Braden, C., 2006. Salmonella enterica Serotype Enteritidis and Eggs: A National Epidemic in the United States. Clin. Infect. Dis. 43, 512–517. https://doi.org/10.1086/505973

Brenner, F.W., Villar, R.G., Angulo, F.J., Tauxe, R., Swaminathan, B., 2000. Salmonella nomenclature. J. Clin. Microbiol. 38, 2465–2467.

Brink, T., Leiss, V., Siegert, P., Jehle, D., Ebner, J.K., Schwan, C., Shymanets, A., Wiese, S., Nürnberg, B., Hensel, M., Aktories, K., Orth, J.H.C., 2018. Salmonella Typhimurium effector SseI inhibits chemotaxis and increases host cell survival by deamidation of heterotrimeric Gi proteins. PLOS Pathog. 14, e1007248.

Briones, V., Tellez, S., Goyache, J., Ballesteros, C., del Pilar Lanzarot, M., Dominguez, L., Fernandez-Garayzabal, J.F., 2004. Salmonella diversity associated with wild reptiles and amphibians in Spain. Environ. Microbiol. 6, 868–871. https://doi.org/10.1111/j.1462-2920.2004.00631.x

Brown, D.F.J., Wootton, M., Howe, R.A., 2015. Antimicrobial susceptibility testing breakpoints and methods from BSAC to EUCAST. J. Antimicrob. Chemother. 71, 3–5. https://doi.org/10.1093/jac/dkv287

Brussow, H., Canchaya, C., Hardt, W.-D., 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol. Mol. Biol. Rev. 68, 560–602. https://doi.org/10.1128/MMBR.68.3.560-602.2004

Brynildsrud, O., Bohlin, J., Scheffer, L., Eldholm, V., 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol. 17, 238. https://doi.org/10.1186/s13059-016-1108-8

Buckle, G.C., Walker, C.L.F., Black, R.E., 2012. Typhoid fever and paratyphoid fever: Systematic review to estimate global morbidity and mortality for 2010. J. Glob. Health 2, 10401. https://doi.org/10.7189/jogh.02.010401

Canals, R., Hammarlof, D.L., Kroger, C., Owen, S. V, Fong, W.Y., Lacharme-Lora, L., Zhu, X., Wenner, N., Carden, S.E., Honeycutt, J., Monack, D.M., Kingsley, R.A., Brownridge, P., Chaudhuri, R.R., Rowe, W.P.M., Predeus, A. V, Hokamp, K., Gordon, M.A., Hinton, J.C.D., 2019. Adding function to the genome of African Salmonella Typhimurium ST313 strain D23580. PLOS Biol. 17, e3000059. https://doi.org/10.1371/journal.pbio.3000059

Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., Møller Aarestrup, F., Hasman, H., 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob. Agents Chemother. 58, 3895–3903. https://doi.org/10.1128/AAC.02412-14

Carden, S.E., Walker, G.T., Honeycutt, J., Lugo, K., Pham, T., Jacobson, A., Bouley, D., Idoyaga, J., Tsolis, R.M., Monack, D., 2017. Pseudogenization of the Secreted Effector Gene sseI Confers Rapid Systemic Dissemination of S. Typhimurium ST313 within Migratory Dendritic Cells. Cell Host Microbe 21, 182–194. https://doi.org/10.1016/j.chom.2017.01.009

Carneiro, M.R.P., Patrício, M.I. de A., Jain, S., Rodrigues, D.D.P., Fracalanzza, S.E.L., 2018. Meningitis caused by Salmonella enterica serotype Panama in Brazil: first case reported. Rev. Soc. Bras. Med. Trop. 51, 244–246. https://doi.org/10.1590/0037-8682-0367-2017

Cartwright, E.J., Jackson, K.A., Johnson, S.D., Graves, L.M., Silk, B.J., Mahon, B.E., 2013. Listeriosis outbreaks and associated food vehicles, United States, 1998-2008. Emerg. Infect. Dis. 19, 1–184. https://doi.org/10.3201/eid1901.120393

Carver, T., Harris, S.R., Berriman, M., Parkhill, J., McQuillan, J.A., 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28, 464—469. https://doi.org/10.1093/bioinformatics/btr703

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.-A., Barrell, B.G., Parkhill, J., 2005. ACT: the Artemis Comparison Tool. Bioinformatics 21, 3422—3423. https://doi.org/10.1093/bioinformatics/bti553

Centers for Disease Control and Prevention, 2018. National Salmonella Surveillance Annual Report, 2016. Atlanta, Georgia.

Chain, P.S.G., Carniel, E., Larimer, F.W., Lamerdin, J., Stoutland, P.O., Regala, W.M., Georgescu, A.M., Vergez, L.M., Land, M.L., Motin, V.L., Brubaker, R.R., Fowler, J., Hinnebusch, J., Marceau, M., Medigue, C., Simonet, M., Chenal-Francisque, V., Souza, B., Dacheux, D., Elliott, J.M., Derbise, A., Hauser, L.J., Garcia, E., 2004. Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis. Proc. Natl. Acad. Sci. U. S. A. 101, 13826–13831. https://doi.org/10.1073/pnas.0404012101

Chaudhuri, R.R., Morgan, E., Peters, S.E., Pleasance, S.J., Hudson, D.L., Davies, H.M., Wang, J., van Diemen, P.M., Buckley, A.M., Bowen, A.J., Pullinger, G.D., Turner, D.J., Langridge, G.C., Turner, A.K., Parkhill, J., Charles, I.G., Maskell, D.J., Stevens, M.P., 2013. Comprehensive assignment of roles for Salmonella Typhimurium genes in intestinal colonization of food-producing animals. PLOS Genet. 9, e1003456. https://doi.org/10.1371/journal.pgen.1003456

Chen, P.E., Shapiro, B.J., 2015. The advent of genome-wide association studies for bacteria. Curr. Opin. Microbiol. 25, 17–24. https://doi.org/10.1016/j.mib.2015.03.002

Cherubin, C.E., 1981. Antibiotic resistance of Salmonella in Europe and the United States. Rev. Infect. Dis. 3, 1105–1126.

Chippaux, J.-P., 2017. Snakebite envenomation turns again into a neglected tropical disease! J. Venom. Anim. Toxins Incl. Trop. Dis. 23, 38. https://doi.org/10.1186/s40409-017-0127-6

Chiu, C.-H., Su, L.-H., Chu, C., 2004. Salmonella enterica serotype Choleraesuis: epidemiology, pathogenesis, clinical disease, and treatment. Clin. Microbiol. Rev. 17, 311–322. https://doi.org/10.1128/cmr.17.2.311-322.2004

Chiu, C.-H., Tang, P., Chu, C., Hu, S., Bao, Q., Yu, J., Chou, Y.-Y., Wang, H.-S., Lee, Y.-S., 2005. The genome sequence of Salmonella enterica serovar Choleraesuis, a highly invasive and resistant zoonotic pathogen. Nucleic Acids Res. 33, 1690–1698. https://doi.org/10.1093/nar/gki297

Clancy, M.M., Davis, M., Valitutto, M.T., Nelson, K., Sykes, J.M. 4th, 2016. Salmonella infection and carriage in reptiles in a zoological collection. J. Am. Vet. Med. Assoc. 248, 1050–1059. https://doi.org/10.2460/javma.248.9.1050

Clarke, J., Wu, H.-C., Jayasinghe, L., Patel, A., Reid, S., Bayley, H., 2009. Continuous base identification for single-molecule nanopore DNA sequencing. Nat. Nanotechnol. 4, 265–270. https://doi.org/10.1038/nnano.2009.12

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honoré, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.-A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J.R., Barrell, B.G., 2001. Massive gene decay in the leprosy bacillus. Nature 409, 1007–1011. https://doi.org/10.1038/35059006

Compeau, P.E.C., Pevzner, P.A., Tesler, G., 2011. How to apply de Bruijn graphs to genome assembly. Nat. Biotechnol. 29, 987–991. https://doi.org/10.1038/nbt.2023

Conway, J.R., Lex, A., Gehlenborg, N., 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33, 2938–2940. https://doi.org/10.1093/bioinformatics/btx364

Cordano, A.M., Virgilio, R., 1996. Evolution of drug resistance in Salmonella panama isolates in Chile. Antimicrob. Agents Chemother. 40, 336–341.

Crosa, J.H., Brenner, D.J., Ewing, W.H., Falkow, S., 1973. Molecular relationships among the Salmonelleae. J. Bacteriol. 115, 307–315. https://doi.org/10.1128/JB.115.1.307-315.1973

Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D., Parkhill, J., Harris, S.R., 2014. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 43, e15–e15. https://doi.org/10.1093/nar/gku1196

Crump, J.A., Luby, S.P., Mintz, E.D., 2004. The global burden of typhoid fever. Bull. World Health Organ. 82, 346–353.

Crump, J.A., Sjölund-Karlsson, M., Gordon, M.A., Parry, C.M., 2015. Epidemiology, Clinical Presentation, Laboratory Diagnosis, Antimicrobial Resistance, and Antimicrobial Management of Invasive Salmonella Infections. Clin. Microbiol. Rev. 28, 901–937. https://doi.org/10.1128/CMR.00002-15

Crump, J.A., Thomas, K.M., Benschop, J., Knox, M.A., Wilkinson, D.A., Midwinter, A.C., Munyua, P., Ochieng, J.B., Bigogo, G.M., Verani, J.R., Widdowson, M.-A., Prinsen, G., Cleaveland, S., Karimuribo, E.D., Kazwala, R.R., Mmbaga, B.T., Swai, E.S., French, N.P., Zadoks, R.N., 2020. Investigating the meat pathway as a source of human nontyphoidal Salmonella bloodstream infections and diarrhea in East Africa. Clin. Infect. Dis. ciaa1153. https://doi.org/10.1093/cid/ciaa1153

Cummins, A.J., Atia, W.A., 1994. Bartholin's abscess complicating food poisoning with Salmonella panama: a case report. Genitourin. Med. 70, 46–48.

Cuypers, W.L., Jacobs, J., Wong, V., Klemm, E.J., Deborggraeve, S., Van Puyvelde, S., 2018. Fluoroquinolone resistance in Salmonella: insights by whole-genome sequencing. Microb. Genomics 4, e000195. https://doi.org/10.1099/mgen.0.000195

Dallman, T., Ashton, P., Schafer, U., Jironkin, A., Painset, A., Shaaban, S., Hartman, H., Myers, R., Underwood, A., Jenkins, C., Grant, K., 2018. SnapperDB: a database solution for routine sequencing analysis of bacterial isolates. Bioinformatics 34, 3028–3029. https://doi.org/10.1093/bioinformatics/bty212

Dallman, T., Inns, T., Jombart, T., Ashton, P., Loman, N., Chatt, C., Messelhaeusser, U., Rabsch, W., Simon, S., Nikisins, S., Bernard, H., le Hello, S., Jourdan da-Silva, N., Kornschober, C., Mossong, J., Hawkey, P., de Pinna, E., Grant, K., Cleary, P., 2016. Phylogenetic structure of European Salmonella Enteritidis outbreak correlates with national and international egg distribution network. Microb. Genomics 2, e000070–

e000070. https://doi.org/10.1099/mgen.0.000070

Daus, M.L., Berendt, S., Wuttge, S., Schneider, E., 2007. Maltose binding protein (MalE) interacts with periplasmic loops P2 and P1 respectively of the MalFG subunits of the maltose ATP binding cassette transporter (MalFGK(2)) from Escherichia coli/Salmonella during the transport cycle. Mol. Microbiol. 66, 1107–1122. https://doi.org/10.1111/j.1365-2958.2007.05982.x

Davis, S., Pettengill, J., Luo, Y., Payne, J., Shpuntoff, A., Rand, H., Strain, E., 2015. CFSAN SNP Pipeline: An automated method for constructing SNP matrices fromnext-generation sequence data. PeerJ Comput. Sci. 1, e20. https://doi.org/10.7717/peerj-cs.20

de Jong, H.K., Parry, C.M., van der Poll, T., Wiersinga, W.J., 2012. Host-pathogen interaction in invasive Salmonellosis. PLOS Pathog. 8, e1002933. https://doi.org/10.1371/journal.ppat.1002933

den Bakker, H.C., Moreno Switt, A.I., Govoni, G., Cummings, C.A., Ranieri, M.L., Degoricija, L., Hoelzer, K., Rodriguez-Rivera, L.D., Brown, S., Bolchacova, E., Furtado, M.R., Wiedmann, M., 2011. Genome sequencing reveals diversification of virulence factor content and possible host adaptation in distinct subpopulations of Salmonella enterica. BMC Genomics 12, 425. https://doi.org/10.1186/1471-2164-12-425

Deng, X., den Bakker, H.C., Hendriksen, R.S., 2016. Genomic Epidemiology: Whole-Genome-Sequencing-Powered Surveillance and Outbreak Investigation of Foodborne Bacterial Pathogens. Annu. Rev. Food Sci. Technol. 7, 353–374. https://doi.org/10.1146/annurev-food-041715-033259

Desai, P.T., Porwollik, S., Long, F., Cheng, P., Wollam, A., Clifton, S.W., Weinstock, G.M., McClelland, M., 2013. Evolutionary Genomics of Salmonella enterica Subspecies. MBio 4, e00579-12. https://doi.org/10.1128/mBio.00579-12

Desjardins, C.A., Cohen, K.A., Munsamy, V., Abeel, T., Maharaj, K., Walker, B.J., Shea, T.P., Almeida, D. V, Manson, A.L., Salazar, A., Padayatchi, N., O'Donnell, M.R., Mlisana, K.P., Wortman, J., Birren, B.W., Grosset, J., Earl, A.M., Pym, A.S., 2016. Genomic and functional analyses of Mycobacterium tuberculosis strains implicate ald in D-cycloserine resistance. Nat. Genet. 48, 544–551. https://doi.org/10.1038/ng.3548

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214. https://doi.org/10.1186/1471-2148-7-214

Edel, W., van Schothorst, M., van Leusden, F.M., Kampelmacher, E.H., 1978. Epidemiological studies on salmonella in a certain area ("Walcheren project") III. The presence of salmonella in man, insects, seagulls and in foods, chopping-block scrapings from butcher's shops, effluent of sewage treatment plants and drains of butcher'. Zentralbl. Bakteriol. Orig. A. 242, 468–480.

EFSA, ECDC, 2018. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2017. EFSA J. 16, 262. https://doi.org/10.2903/j.efsa.2018.5500

Eisenstein, M., 2012. Oxford Nanopore announcement sets sequencing sector abuzz. Nat. Biotechnol. https://doi.org/10.1038/nbt0412-295

Elenga, N., Cuadro, E., Long, L., Njuieyon, F., Martin, E., Kom-Tchameni, R., Defo, A., Razafindrakoto, S.H., Mrsic, Y., Henaff, F., Mahamat, A., 2017. Salmonella enterica serovar Panama meningitis in exclusive breastfeeding infants: Report of 4 cases, clinical features and therapeutic challenges. Medicine (Baltimore). 96, e6665. https://doi.org/10.1097/MD.0000000000006665

Emond-Rheault, J.-G., Hamel, J., Jeukens, J., Freschi, L., Kukavica-Ibrulj, I., Boyle, B., Tamber, S., Malo, D., Franz, E., Burnett, E., Daigle, F., Arya, G., Sanderson, K., Wiedmann, M., Slawson, R.M., Weadge, J.T., Stephan, R., Bekal, S., Gruenheid, S., Goodridge, L.D., Levesque, R.C., 2020. The Salmonella enterica Plasmidome as a Reservoir of Antibiotic Resistance. Microorganisms 8, 1016. https://doi.org/10.3390/microorganisms8071016

Eng, S.-K., Pusparajah, P., Ab Mutalib, N.-S., Ser, H.-L., Chan, K.-G., Lee, L.-H., 2015. Salmonella: A review on pathogenesis, epidemiology and antibiotic resistance. Front. Life Sci. 8, 284–293. https://doi.org/10.1080/21553769.2015.1051243

Ernst, R., Gurdan, P., 1973. [Salmonella-panama epidemic in Basel, spring 1972, from the veterinary

viewpoint]. Schweiz. Arch. Tierheilkd. 115, 8–15.

European Food Safety Authority and European Centre for Disease Prevention and Control, 2012. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2010. Euro Surveill. Eur. Commun. Dis. Bull.

European Food Safety Authority and European Centre for Disease Prevention and Control (EFSA and ECDC), 2019. The European Union One Health 2018 Zoonoses Report. EFSA J. 17, e05926. https://doi.org/10.2903/j.efsa.2019.5926

Falush, D., Torpdahl, M., Didelot, X., Conrad, D.F., Wilson, D.J., Achtman, M., 2006. Mismatch induced speciation in Salmonella: model and data. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 361, 2045–2053. https://doi.org/10.1098/rstb.2006.1925

Fan, H., Chu, J.-Y., 2007. A brief review of short tandem repeat mutation. Genomics. Proteomics Bioinformatics 5, 7–14. https://doi.org/10.1016/S1672-0229(07)60009-6

Fang, Y., Mercer, R.G., McMullen, L.M., Gänzle, M.G., 2017. Induction of Shiga Toxin-Encoding Prophage by Abiotic Environmental Stress in Food. Appl. Environ. Microbiol. 83. https://doi.org/10.1128/AEM.01378-17

Feasey, N.A., Cain, A.K., Msefula, C.L., Pickard, D., Alaerts, M., Aslett, M., Everett, D.B., Allain, T.J., Dougan, G., Gordon, M.A., Heyderman, R.S., Kingsley, R.A., 2014. Drug Resistance in Salmonella enterica ser. Typhimurium Bloodstream Infection, Malawi. Emerg. Infect. Dis. 20, 1957–1959. https://doi.org/10.3201/eid2011.141175

Feasey, N.A., Dougan, G., Kingsley, R.A., Heyderman, R.S., Gordon, M.A., 2012. Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. Lancet 379, 2489–2499. https://doi.org/10.1016/S0140-6736(11)61752-2

Feasey, N.A., Hadfield, J., Keddy, K.H., Dallman, T.J., Jacobs, J., Deng, X., Wigley, P., Barquist, L., Langridge, G.C., Feltwell, T., Harris, S.R., Mather, A.E., Fookes, M., Aslett, M., Msefula, C., Kariuki, S., Maclennan, C.A., Onsare, R.S., Weill, F.-X., Le Hello, S., Smith, A.M., McClelland, M., Desai, P., Parry, C.M., Cheesbrough, J., French, N., Campos, J., Chabalgoity, J.A., Betancor, L., Hopkins, K.L., Nair, S., Humphrey, T.J., Lunguya, O., Cogan, T.A., Tapia, M.D., Sow, S.O., Tennant, S.M., Bornstein, K., Levine, M.M., Lacharme-Lora, L., Everett, D.B., Kingsley, R.A., Parkhill, J., Heyderman, R.S., Dougan, G., Gordon, M.A., Thomson, N.R., 2016. Distinct Salmonella Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. Nat. Genet. 48, 1211–1217. https://doi.org/10.1038/ng.3644

Feasey, N.A., Masesa, C., Jassi, C., Faragher, E.B., Mallewa, J., Mallewa, M., MacLennan, C.A., Msefula, C., Heyderman, R.S., Gordon, M.A., 2015. Three Epidemics of Invasive Multidrug-Resistant Salmonella Bloodstream Infection in Blantyre, Malawi, 1998-2014. Clin. Infect. Dis. 61 Suppl 4, S363-71. https://doi.org/10.1093/cid/civ691

Feldgarden, M., Brover, V., Haft, D.H., Prasad, A.B., Slotta, D.J., Tolstoy, I., Tyson, G.H., Zhao, S., Hsu, C.-H., McDermott, P.F., Tadesse, D.A., Morales, C., Simmons, M., Tillman, G., Wasilenko, J., Folster, J.P., Klimke, W., 2019. Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. Antimicrob. Agents Chemother. 63, e00483-19. https://doi.org/10.1128/AAC.00483-19

Feldheim, D.A., Chin, A.M., Nierva, C.T., Feucht, B.U., Cao, Y.W., Xu, Y.F., Sutrina, S.L., Saier, M.H.J., 1990. Physiological consequences of the complete loss of phosphoryl-transfer proteins HPr and FPr of the phosphoenolpyruvate:sugar phosphotransferase system and analysis of fructose (fru) operon expression in Salmonella typhimurium. J. Bacteriol. 172, 5459–5469.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17, 368–376. https://doi.org/10.1007/BF01734359

Feng, Y., Johnston, R.N., Liu, G.-R., Liu, S.-L., 2013. Genomic Comparison between Salmonella Gallinarum and Pullorum: Differential Pseudogene Formation under Common Host Restriction. PLoS One 8, e59427.

Ferragina, P., Manzini, G., 2000. Opportunistic data structures with applications, in: Proceedings 41st Annual Symposium on Foundations of Computer Science. pp. 390–398. https://doi.org/10.1109/SFCS.2000.892127

Ferrari, R.G., Rosario, D.K.A., Cunha-Neto, A., Mano, S.B., Figueiredo, E.E.S., Conte-Junior, C.A., 2019. Worldwide Epidemiology of Salmonella Serovars in Animal-Based Foods: a Meta-analysis. Appl. Environ. Microbiol. 85, e00591-19. https://doi.org/10.1128/AEM.00591-19

Figueroa-Bossi, N., Uzzau, S., Maloriol, D., Bossi, L., 2001. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in Salmonella. Mol. Microbiol. 39, 260–271. https://doi.org/10.1046/j.1365-2958.2001.02234.x

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., al., et, 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science (80-.). 269, 496 LP – 512. https://doi.org/10.1126/science.7542800

Fookes, M., Schroeder, G.N., Langridge, G.C., Blondel, C.J., Mammina, C., Connor, T.R., Seth-Smith, H., Vernikos, G.S., Robinson, K.S., Sanders, M., Petty, N.K., Kingsley, R.A., Bäumler, A.J., Nuccio, S.-P., Contreras, I., Santiviago, C.A., Maskell, D., Barrow, P., Humphrey, T., Nastasi, A., Roberts, M., Frankel, G., Parkhill, J., Dougan, G., Thomson, N.R., 2011. Salmonella bongori Provides Insights into the Evolution of the Salmonellae. PLOS Pathog. 7, e1002191.

Fox, D.K., Roseman, S., 1986. Isolation and characterization of homogeneous acetate kinase from Salmonella typhimurium and Escherichia coli. J. Biol. Chem. 261, 13487–13497.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, R.D., Weidman, J.F., Small, K. V, Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A. 3rd, Venter, J.C., 1995. The minimal gene complement of Mycoplasma genitalium. Science 270, 397–403. https://doi.org/10.1126/science.270.5235.397

Fuller, C.C., Jawahir, S.L., Leano, F.T., Bidol, S.A., Signs, K., Davis, C., Holmes, Y., Morgan, J., Teltow, G., Jones, B., Sexton, R.B., Davis, G.L., Braden, C.R., Patel, N.J., Deasy, M.P. 3rd, Smith, K.E., 2008. A multi-state Salmonella Typhimurium outbreak associated with frozen vacuum-packed rodents used to feed snakes. Zoonoses Public Health 55, 481–487.

Gaffga, N.H., Barton Behravesh, C., Ettestad, P.J., Smelser, C.B., Rhorer, A.R., Cronquist, A.B., Comstock, N.A., Bidol, S.A., Patel, N.J., Gerner-Smidt, P., Keene, W.E., Gomez, T.M., Hopkins, B.A., Sotir, M.J., Angulo, F.J., 2012. Outbreak of salmonellosis linked to live poultry from a mail-order hatchery. N. Engl. J. Med. 366, 2065–2073. https://doi.org/10.1056/NEJMoa1111818

Gal-Mor, O., Boyle, E.C., Grassl, G.A., 2014. Same species, different diseases: how and why typhoidal and non-typhoidal Salmonella enterica serovars differ. Front. Microbiol. 5, 391. https://doi.org/10.3389/fmicb.2014.00391

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T.F., Conesa, A., 2012. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics 28, 2678–2679. https://doi.org/10.1093/bioinformatics/bts503

Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. arXiv 1207.

Gast, R.K., Guraya, R., Jones, D.R., Anderson, K.E., Karcher, D.M., 2017. Frequency and Duration of Fecal Shedding of Salmonella Enteritidis by Experimentally Infected Laying Hens Housed in Enriched Colony Cages at Different Stocking Densities. Front. Vet. Sci. 4, 47. https://doi.org/10.3389/fvets.2017.00047

Gay, N., Le Hello, S., Weill, F.-X., de Thoisy, B., Berger, F., 2014. Salmonella serotypes in reptiles and humans, French Guiana. Vet. Microbiol. 170, 167–171. https://doi.org/10.1016/j.vetmic.2014.01.024

Gericke, D., Luchtrath, H., 1951. [A peculiar form of meningitis caused by Salmonella panama]. Med. Klin. 46, 862–865.

Geue, L., Loschner, U., 2002. Salmonella enterica in reptiles of German and Austrian origin. Vet. Microbiol. 84, 79–91.

Gibani, M.M., Jones, E., Barton, A., Jin, C., Meek, J., Camara, S., Galal, U., Heinz, E., Rosenberg-Hasson, Y., Obermoser, G., Jones, C., Campbell, D., Black, C., Thomaides-Brears, H., Darlow, C., Dold, C., Silva-Reyes, L., Blackwell, L., Lara-Tejero, M., Jiao, X., Stack, G., Blohmke, C.J., Hill, J., Angus, B., Dougan, G., Galán, J.,

Pollard, A.J., 2019. Investigation of the role of typhoid toxin in acute typhoid fever in a human challenge model. Nat. Med. 25, 1082–1088. https://doi.org/10.1038/s41591-019-0505-4

Gibbons, H.S., Lin, S., Cotter, R.J., Raetz, C.R., 2000. Oxygen requirement for the biosynthesis of the S-2-hydroxymyristate moiety in Salmonella typhimurium lipid A. Function of LpxO, A new Fe2+/alpha-ketoglutarate-dependent dioxygenase homologue. J. Biol. Chem. 275, 32940–32949. https://doi.org/10.1074/jbc.M005779200

Gieraltowski, L., Julian, E., Pringle, J., Macdonald, K., Quilliam, D., Marsden-Haug, N., Saathoff-Huber, L., Von Stein, D., Kissler, B., Parish, M., Elder, D., Howard-King, V., Besser, J., Sodha, S., Loharikar, A., Dalton, S., Williams, I., Barton Behravesh, C., 2013. Nationwide outbreak of Salmonella Montevideo infections associated with contaminated imported black and red pepper: warehouse membership cards provide critical clues to identify the source. Epidemiol. Infect. 141, 1244–1252. https://doi.org/10.1017/S0950268812001859

Gilchrist, J.J., MacLennan, C.A., 2019. Invasive Nontyphoidal Salmonella Disease in Africa. EcoSal Plus 8. https://doi.org/10.1128/ecosalplus.ESP-0007-2018

Gilks, C.F., Brindle, R.J., Otieno, L.S., Simani, P.M., Newnham, R.S., Bhatt, S.M., Lule, G.N., Okelo, G.B., Watkins, W.M., Waiyaki, P.G., 1990. Life-threatening bacteraemia in HIV-1 seropositive adults admitted to hospital in Nairobi, Kenya. Lancet 336, 545–549. https://doi.org/10.1016/0140-6736(90)92096-z

Giribet, G., 2007. Efficient tree searches with available algorithms. Evol. Bioinform. Online 3, 341–356.

Gordon, M.A., 2008. Salmonella infections in immunocompromised adults. J. Infect. 56, 413–422. https://doi.org/10.1016/j.jinf.2008.03.012

Goupil, B.A., Trent, A.M., Bender, Jeff, B., Olsen, Karen, E., Morningstar, B.R., Wuenschmann, A., 2012. A Longitudinal Study of Salmonella from Snakes Used in a Public Outreach Program. J. Zoo Wildl. Med. 43, 836–841.

Graham, R.M.A., Hiley, L., Rathnayake, I.U., Jennison, A. V., 2018. Comparative genomics identifies distinct lineages of S. Enteritidis from Queensland, Australia. PLoS One 13, 1–15. https://doi.org/10.1371/journal.pone.0191042

Grimont, P., Weill, F.-X., 2007. Antigenic Formulae of the Salmonella serovars, (9th ed.) Paris: WHO Collaborating Centre for Reference and Research on Salmonella. Inst. Pasteur. 1–166.

Guinee, P.A., 1969. Phage types and resistance factors in S. panama strains from various countries. Zentralbl. Bakteriol. Orig. 209, 331–336.

Guinee, P.A., 1968. R transfer to S. panama in vitro and in vivo. Antonie Van Leeuwenhoek 34, 93–98.

Guinee, P.A., Scholtens, R.T., Willems, H.M., 1967. Influence of resistance-factors on the phage types of Salmonella panama. Antonie Van Leeuwenhoek 33, 30–40.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Gutnick, D., Calvo, J.M., Klopotowski, T., Ames, B.N., 1969. Compounds which serve as the sole source of carbon or nitrogen for Salmonella typhimurium LT-2. J. Bacteriol. 100, 215–219.

Guyomard-Rabenirina, S., Muanza, B., Bastian, S., Malpote, E., Jestin, P., Guerin, M., Talarmin, A., Weill, F.-X., Legrand, A., Breurec, S., 2018. Salmonella enterica serovars Panama and Arechavaleta: Risk Factors for Invasive Non-Typhoidal Salmonella Disease in Guadeloupe, French West Indies. Am. J. Trop. Med. Hyg. 99, 584–589. https://doi.org/10.4269/ajtmh.18-0192

Guyomard-Rabenirina, S., Weill, F.-X., Le Hello, S., Bastian, S., Berger, F., Ferdinand, S., Legreneur, P., Loraux, C., Malpote, E., Muanza, B., Richard, V., Talarmin, A., Breurec, S., 2019. Reptiles in Guadeloupe (French West Indies) are a reservoir of major human Salmonella enterica serovars. PLoS One 14, e0220145.

Hammarlöf, D.L., Kröger, C., Owen, S. V, Canals, R., Lacharme-Lora, L., Wenner, N., Schager, A.E., Wells, T.J., Henderson, I.R., Wigley, P., Hokamp, K., Feasey, N.A., Gordon, M.A., Hinton, J.C.D., 2018. Role of a single noncoding nucleotide in the evolution of an epidemic African clade of Salmonella. Proc. Natl. Acad. Sci. 115, E2614–E2623. https://doi.org/10.1073/pnas.1714718115

Hanna, L.F., Matthews, T.D., Dinsdale, E.A., Hasty, D., Edwards, R.A., 2012. Characterization of the ELPhiS Prophage from Salmonella enterica Serovar Enteritidis Strain LK5. Appl. Environ. Microbiol. 78, 1785–1793. https://doi.org/10.1128/AEM.07241-11

Harrison, R.A., Hargreaves, A., Wagstaff, S.C., Faragher, B., Lalloo, D.G., 2009. Snake Envenoming: A Disease of Poverty. PLoS Negl. Trop. Dis. 3, e569. https://doi.org/10.1371/journal.pntd.0000569

Hatem, A., Bozdağ, D., Toland, A.E., Çatalyürek, Ü. V, 2013. Benchmarking short sequence mapping tools. BMC Bioinformatics 14, 184. https://doi.org/10.1186/1471-2105-14-184

Hawken, S.E., Snitkin, E.S., 2019. Genomic epidemiology of multidrug-resistant Gram-negative organisms. Ann. N. Y. Acad. Sci. 1435, 39–56. https://doi.org/10.1111/nyas.13672

Heather, J.M., Chain, B., 2016. The sequence of sequencers: The history of sequencing DNA. Genomics 107, 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Hedge, J., Wilson, D.J., 2014. Bacterial phylogenetic reconstruction from whole genomes is robust to recombination but demographic inference is not. MBio 5, e02158. https://doi.org/10.1128/mBio.02158-14

Hendriksen, R.S., Bortolaia, V., Tate, H., Tyson, G.H., Aarestrup, F.M., McDermott, P.F., 2019. Using Genomics to Track Global Antimicrobial Resistance . Front. Public Heal. .

Hendriksen, R.S., Vieira, A.R., Karlsmose, S., Lo Fo Wong, D.M.A., Jensen, A.B., Wegener, H.C., Aarestrup, F.M., 2011. Global monitoring of Salmonella serovar distribution from the World Health Organization Global Foodborne Infections Network Country Data Bank: results of quality assured laboratories from 2001 to 2007. Foodborne Pathog. Dis. 8, 887–900. https://doi.org/10.1089/fpd.2010.0787

Herrero-Fresno, A., Wallrodt, I., Leekitcharoenphon, P., Olsen, J.E., Aarestrup, F.M., Hendriksen, R.S., 2014. The role of the st313-td gene in virulence of Salmonella Typhimurium ST313. PLoS One 9, e84566. https://doi.org/10.1371/journal.pone.0084566

Heysell, S.K., Thomas, T.A., Morrison, A.R., Barry, M., 2008. Salmonella panama and acute respiratory distress syndrome in a traveler taking a proton pump inhibitor. J. Travel Med. 15, 460–463. https://doi.org/10.1111/j.1708-8305.2008.00258.x

Ho, T.D., Slauch, J.M., 2001. Characterization of grvA, an antivirulence gene on the gifsy-2 phage in Salmonella enterica serovar typhimurium. J. Bacteriol. 183, 611–620. https://doi.org/10.1128/JB.183.2.611-620.2001

Hoelzer, K., Moreno Switt, A.I., Wiedmann, M., 2011. Animal contact as a source of human non-typhoidal salmonellosis. Vet. Res. 42, 34. https://doi.org/10.1186/1297-9716-42-34

Holt, K.E., Baker, S., Weill, F.-X., Holmes, E.C., Kitchen, A., Yu, J., Sangal, V., Brown, D.J., Coia, J.E., Kim, D.W., Choi, S.Y., Kim, S.H., da Silveira, W.D., Pickard, D.J., Farrar, J.J., Parkhill, J., Dougan, G., Thomson, N.R., 2012. Shigella sonnei genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. Nat. Genet. 44, 1056–1059. https://doi.org/10.1038/ng.2369

Holt, K.E., Thomson, N.R., Wain, J., Langridge, G.C., Hasan, R., Bhutta, Z.A., Quail, M.A., Norbertczak, H., Walker, D., Simmonds, M., White, B., Bason, N., Mungall, K., Dougan, G., Parkhill, J., 2009. Pseudogene accumulation in the evolutionary histories of Salmonella enterica serovars Paratyphi A and Typhi. BMC Genomics 10, 36. https://doi.org/10.1186/1471-2164-10-36

Honeycutt, J.D., Wenner, N., Li, Y., Brewer, S.M., Massis, L.M., Brubaker, S.W., Chairatana, P., Owen, S. V, Canals, R., Hinton, J.C.D., Monack, D.M., 2020. Genetic variation in the MacAB-TolC efflux pump influences pathogenesis of invasive Salmonella isolates from Africa. PLOS Pathog. 16, e1008763. https://doi.org/10.1371/journal.ppat.1008763

Hooper, D.C., Jacoby, G.A., 2015. Mechanisms of drug resistance: quinolone resistance. Ann. N. Y. Acad. Sci. 1354, 12–31. https://doi.org/10.1111/nyas.12830

Hopkins, K.L., Kirchner, M., Guerra, B., Granier, S.A., Lucarelli, C., Porrero, M.C., Jakubczak, A., Threlfall, E.J., Mevius, D.J., 2010. Multiresistant Salmonella enterica serovar 4,[5],12:i:- in Europe: a new pandemic strain? Euro Surveill. Eur. Commun. Dis. Bull. 15, 19580.

Horiuchi, S., Inagaki, Y., Nakaya, R., Goto, N., Yoshida, Y., Kusunoki, J., Ito, T., Ohashi, M., 1989. [Serovars, antimicrobial resistance and conjugative R plasmids of Salmonella isolated from human during the period of 1966-1986 in Tokyo. Research Group for Infectious Enteric Diseases]. Kansenshogaku Zasshi. 63, 352–362.

Huang, S.-C., Chiu, C.-H., Chiou, C.-S., Yang, Y.-J., 2013. Multidrug-resistant Salmonella enterica serovar Panama carrying class 1 integrons is invasive in Taiwanese children. J. Formos. Med. Assoc. 112, 269–275. https://doi.org/10.1016/j.jfma.2012.02.011

Hunkapiller, T., Kaiser, R.J., Koop, B.F., Hood, L., 1991. Large-scale and automated DNA sequence determination. Science (80-.). 254, 59–67. https://doi.org/10.1126/science.1925562

Hunt, M., Mather, A.E., Sánchez-Busó, L., Page, A.J., Parkhill, J., Keane, J.A., Harris, S.R., 2017. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb. Genomics 3, e000131–e000131. https://doi.org/10.1099/mgen.0.000131

Hyman, E.D., 1988. A new method of sequencing DNA. Anal. Biochem. 174, 423–436. https://doi.org/10.1016/0003-2697(88)90041-3

Ibrahim, A.A., 1969. Bacteriophage typing of Salmonella. I. Isolation and host range study of bacteriophages. Appl. Microbiol. 18, 444–447.

Ingle, D.J., Nair, S., Hartman, H., Ashton, P.M., Dyson, Z.A., Day, M., Freedman, J., Chattaway, M.A., Holt, K.E., Dallman, T.J., 2019. Informal genomic surveillance of regional distribution of Salmonella Typhi genotypes and antimicrobial resistance via returning travellers. PLoS Negl. Trop. Dis. 13, e0007620.

Inouye, M., Dashnow, H., Raven, L.-A., Schultz, M.B., Pope, B.J., Tomita, T., Zobel, J., Holt, K.E., 2014. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 6, 90. https://doi.org/10.1186/s13073-014-0090-6

Iqbal, Z., Caccamo, M., Turner, I., Flicek, P., McVean, G., 2012. De novo assembly and genotyping of variants using colored de Bruijn graphs. Nat. Genet. 44, 226–232. https://doi.org/10.1038/ng.1028

Issenhuth-Jeanjean, S., Roggentin, P., Mikoleit, M., Guibourdenche, M., de Pinna, E., Nair, S., Fields, P.I., Weill, F.-X., 2014. Supplement 2008-2010 (no. 48) to the White-Kauffmann-Le Minor scheme. Res. Microbiol. 165, 526–530. https://doi.org/10.1016/j.resmic.2014.07.004

Jackman, S.D., Vandervalk, B.P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S.A., Jahesh, G., Khan, H., Coombe, L., Warren, R.L., Birol, I., 2017. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome Res. 27, 768–777. https://doi.org/10.1101/gr.214346.116

Jacob, J.J., Anandan, S., Venkatesan, M., Neeravi, A., Vasudevan, K., Pragasam, A.K., Veeraraghavan, B., 2019. Genomic analysis of human invasive Salmonella enterica serovar Typhimurium ST313 isolate B3589 from India. Infect. Genet. Evol. 73, 416–424. https://doi.org/10.1016/j.meegid.2019.05.023

Jain, M., Olsen, H.E., Paten, B., Akeson, M., 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. 17, 239. https://doi.org/10.1186/s13059-016-1103-0

Jones, T.F., Ingram, L.A., Cieslak, P.R., Vugia, D.J., Tobin-D'Angelo, M., Hurd, S., Medus, C., Cronquist, A., Angulo, F.J., 2008. Salmonellosis outcomes differ substantially by serotype. J. Infect. Dis. 198, 109–114. https://doi.org/10.1086/588823

Kagambega, A., Lienemann, T., Frye, J.G., Barro, N., Haukka, K., 2018. Whole genome sequencing of multidrug-resistant Salmonella enterica serovar Typhimurium isolated from humans and poultry in Burkina Faso. Trop. Med. Health 46, 4. https://doi.org/10.1186/s41182-018-0086-9

Kan, B., Zhou, H., Du, P., Zhang, W., Lu, X., Qin, T., Xu, J., 2018. Transforming bacterial disease surveillance and investigation using whole-genome sequence to probe the trace. Front. Med. 12, 23–33. https://doi.org/10.1007/s11684-017-0607-7

Kanagarajah, S., Waldram, A., Dolan, G., Jenkins, C., Ashton, P.M., Carrion Martin, A.I., Davies, R., Frost, A., Dallman, T.J., De Pinna, E.M., Hawker, J.I., Grant, K.A., Elson, R., 2018. Whole genome sequencing reveals an outbreak of Salmonella Enteritidis associated with reptile feeder mice in the United Kingdom, 2012-2015. Food Microbiol. 71, 32–38. https://doi.org/10.1016/j.fm.2017.04.005

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2014. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 42, D199-205. https://doi.org/10.1093/nar/gkt1076

Kariuki, S., Okoro, C., Kiiru, J., Njoroge, S., Omuse, G., Langridge, G., Kingsley, R.A., Dougan, G., Revathi, G., 2015. Ceftriaxone-resistant Salmonella enterica serotype typhimurium sequence type 313 from Kenyan patients is associated with the blaCTX-M-15 gene on a novel IncHI2 plasmid. Antimicrob. Agents Chemother. 59, 3133–3139. https://doi.org/10.1128/AAC.00078-15

Kariuki, S., Revathi, G., Gakuya, F., Yamo, V., Muyodi, J., Hart, C.A., 2002. Lack of clonal relationship between non-typhi Salmonella strain types from humans and those isolated from animals living in close contact. FEMS Immunol. Med. Microbiol. 33, 165–171. https://doi.org/10.1111/j.1574-695X.2002.tb00587.x

Kariuki, S., Revathi, G., Kariuki, N., Kiiru, J., Mwituria, J., Muyodi, J., Githinji, J.W., Kagendo, D., Munyalo, A., Hart, C.A., 2006. Invasive multidrug-resistant non-typhoidal Salmonella infections in Africa: Zoonotic or anthroponotic transmission? J. Med. Microbiol. 55, 585–591. https://doi.org/10.1099/jmm.0.46375-0

Kauffmann, F., 1966. The bacteriology of Enterobacteriaceae. Copenhagen, Denmark:Munksgaard.

Kauffmann, F., Edwards, P.R., 1952. Classification and Nomenclature of Enterobacteriaceae. Int. Bull. Bacteriol. Nomencl. Taxon. 2, 2–8.

Kempf, G., Pietzsch, O., 1977. [Phage-typing and tetracycline resistance in Salmonella panama strains of animal origin in the Federal Republic of Germany (1969-1975) (author's transl)]. Zentralbl. Bakteriol. Orig. A. 238, 370–378.

Key, F.M., Posth, C., Esquivel-Gomez, L.R., Hübler, R., Spyrou, M.A., Neumann, G.U., Furtwängler, A., Sabin, S., Burri, M., Wissgott, A., Lankapalli, A.K., Vågene, Å.J., Meyer, M., Nagel, S., Tukhbatova, R., Khokhlov, A., Chizhevsky, A., Hansen, S., Belinsky, A.B., Kalmykov, A., Kantorovich, A.R., Maslov, V.E., Stockhammer, P.W., Vai, S., Zavattaro, M., Riga, A., Caramelli, D., Skeates, R., Beckett, J., Gradoli, M.G., Steuri, N., Hafner, A., Ramstein, M., Siebke, I., Lösch, S., Erdal, Y.S., Alikhan, N.-F., Zhou, Z., Achtman, M., Bos, K., Reinhold, S., Haak, W., Kühnert, D., Herbig, A., Krause, J., 2020. Emergence of human-adapted Salmonella enterica is linked to the Neolithization process. Nat. Ecol. Evol. 4, 324–333. https://doi.org/10.1038/s41559-020-1106-9

Kidgell, C., Reichard, U., Wain, J., Linz, B., Torpdahl, M., Dougan, G., Achtman, M., 2002. Salmonella typhi, the causative agent of typhoid fever, is approximately 50,000 years old. Infect. Genet. Evol. 2, 39–45. https://doi.org/https://doi.org/10.1016/S1567-1348(02)00089-8

Kingsley, R.A., Bäumler, A.J., 2000. Host adaptation and the emergence of infectious disease: the Salmonella paradigm. Mol. Microbiol. 36, 1006–1014. https://doi.org/10.1046/j.1365-2958.2000.01907.x

Kingsley, R.A., Msefula, C.L., Thomson, N.R., Kariuki, S., Holt, K.E., Gordon, M.A., Harris, D., Clarke, L., Whitehead, S., Sangal, V., Marsh, K., Achtman, M., Molyneux, M.E., Cormican, M., Parkhill, J., MacLennan, C.A., Heyderman, R.S., Dougan, G., 2009. Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. Genome Res. 19, 2279–2287. https://doi.org/10.1101/gr.091017.109

Kirk, M.D., Pires, S.M., Black, R.E., Caipo, M., Crump, J.A., Devleesschauwer, B., Döpfer, D., Fazil, A., Fischer-Walker, C.L., Hald, T., Hall, A.J., Keddy, K.H., Lake, R.J., Lanata, C.F., Torgerson, P.R., Havelaar, A.H., Angulo, F.J., 2015. World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. PLOS Med. 12, e1001921.

Klemm, E., Dougan, G., 2016. Advances in Understanding Bacterial Pathogenesis Gained from Whole-Genome Sequencing and Phylogenetics. Cell Host Microbe 19, 599–610. https://doi.org/https://doi.org/10.1016/j.chom.2016.04.015

Klemm, E.J., Gkrania-Klotsas, E., Hadfield, J., Forbester, J.L., Harris, S.R., Hale, C., Heath, J.N., Wileman, T., Clare, S., Kane, L., Goulding, D., Otto, T.D., Kay, S., Doffinger, R., Cooke, F.J., Carmichael, A., Lever, A.M.L., Parkhill, J., MacLennan, C.A., Kumararatne, D., Dougan, G., Kingsley, R.A., 2016. Emergence of host-adapted Salmonella Enteritidis through rapid evolution in an immunocompromised host. Nat. Microbiol. 1, 15023. https://doi.org/10.1038/nmicrobiol.2015.23

Klemm, E.J., Shakoor, S., Page, A.J., Qamar, F.N., Judge, K., Saeed, D.K., Wong, V.K., Dallman, T.J., Nair, S.,

Baker, S., Shaheen, G., Qureshi, S., Yousafzai, M.T., Saleem, M.K., Hasan, Z., Dougan, G., Hasan, R., 2018. Emergence of an Extensively Drug-Resistant Salmonella enterica Serovar Typhi Clone Harboring a Promiscuous Plasmid Encoding Resistance to Fluoroquinolones and Third-Generation Cephalosporins. MBio 9, e00105-18. https://doi.org/10.1128/mBio.00105-18

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K., Ding, L., 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25, 2283–2285. https://doi.org/10.1093/bioinformatics/btp373

Kostiala, A.A., Ranta, T., 1989. Pelvic inflammatory disease caused by Salmonella panama and its treatment with ciprofloxacin. Case report. Br. J. Obstet. Gynaecol. 96, 120–122.

Kostiala, A.A., Westerstrahle, M., Muttilainen, M., 1992. Neonatal Salmonella panama infection with meningitis. Acta Paediatr. 81, 856–858.

Kröger, C., Fuchs, T.M., 2009. Characterization of the myo-inositol utilization island of Salmonella enterica serovar Typhimurium. J. Bacteriol. 191, 545–554. https://doi.org/10.1128/JB.01253-08

Krueger, A.L., Greene, S.A., Barzilay, E.J., Henao, O., Vugia, D., Hanna, S., Meyer, S., Smith, K., Pecic, G., Hoefer, D., Griffin, P.M., 2014. Clinical Outcomes of Nalidixic Acid, Ceftriaxone, and Multidrug-Resistant Nontyphoidal Salmonella Infections Compared with Pansusceptible Infections in FoodNet Sites, 2006–2008. Foodborne Pathog. Dis. 11, 335–341. https://doi.org/10.1089/fpd.2013.1642

Kudirkiene, E., Andoh, L.A., Ahmed, S., Herrero-Fresno, A., Dalsgaard, A., Obiri-Danso, K., Olsen, J.E., 2018. The Use of a Combined Bioinformatics Approach to Locate Antibiotic Resistance Genes on Plasmids From Whole Genome Sequences of Salmonella enterica Serovars From Humans in Ghana. Front. Microbiol.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096

Lacharme-Lora, L., Owen, S. V, Blundell, R., Canals, R., Wenner, N., Perez-Sepulveda, B., Fong, W.Y., Gilroy, R., Wigley, P., Hinton, J.C.D., 2019. The use of chicken and insect infection models to assess the virulence of African Salmonella Typhimurium ST313. PLoS Negl. Trop. Dis. 13, e0007540.

Lamanna, A., Benci, G., Alecci, A., 1977. [Sensitivity to chemoantibiotic therapy of new isolated Salmonellae strains (author's transl)]. Ann. Sclavo. 19, 409–419.

Lamas, A., Miranda, J.M., Regal, P., Vázquez, B., Franco, C.M., Cepeda, A., 2018. A comprehensive review of non-enterica subspecies of Salmonella enterica. Microbiol. Res. 206, 60–73. https://doi.org/https://doi.org/10.1016/j.micres.2017.09.010

Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M.R., Ahn, T.-H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S., Ussery, D.W., 2015. Insights from 20 years of bacterial genome sequencing. Funct. Integr. Genomics 15, 141–161. https://doi.org/10.1007/s10142-015-0433-4

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Langridge, G.C., Fookes, M., Connor, T.R., Feltwell, T., Feasey, N., Parsons, B.N., Seth-Smith, H.M.B., Barquist, L., Stedman, A., Humphrey, T., Wigley, P., Peters, S.E., Maskell, D.J., Corander, J., Chabalgoity, J.A., Barrow, P., Parkhill, J., Dougan, G., Thomson, N.R., 2015. Patterns of genome evolution that have accompanied host adaptation in Salmonella. Proc. Natl. Acad. Sci. U. S. A. 112, 863–868. https://doi.org/10.1073/pnas.1416707112

Langridge, G.C., Nair, S., Wain, J., 2009a. Nontyphoidal Salmonella serovars cause different degrees of invasive disease globally. J. Infect. Dis. https://doi.org/10.1086/596208

Langridge, G.C., Phan, M.-D., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., Charles, I., Maskell, D.J., Peters, S.E., Dougan, G., Wain, J., Parkhill, J., Turner, A.K., 2009b. Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome Res. 19, 2308–2316. https://doi.org/10.1101/gr.097097.109

Langridge, G.C., Wain, J., Nair, S., 2012. Invasive Salmonellosis in Humans. EcoSal Plus 5. https://doi.org/10.1128/ecosalplus.8.6.2.2

Lantos, J., Fekete, J., Kiraly, K., 1981. R-plasmid study of an outbreak caused by multiresistant strains of Salmonella panama. Acta Microbiol. Acad. Sci. Hung. 28, 211–217.

Larsen, M. V, Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L., Jelsbak, L., Sicheritz-Ponten, T., Ussery, D.W., Aarestrup, F.M., Lund, O., 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J. Clin. Microbiol. 50, 1355–1361. https://doi.org/10.1128/JCM.06094-11

Lau, M.S.Y., Marion, G., Streftaris, G., Gibson, G., 2015. A Systematic Bayesian Integration of Epidemiological and Genetic Data. PLOS Comput. Biol. 11, e1004633.

Lawley, T.D., Chan, K., Thompson, L.J., Kim, C.C., Govoni, G.R., Monack, D.M., 2006. Genome-Wide Screen for Salmonella Genes Required for Long-Term Systemic Infection of the Mouse. PLOS Pathog. 2, e11.

Lazarescu, M., Mucuta, G., Pena, J., 1971. [Meningitis with Salmonella panama in infants]. Microbiol. Parazitol. Epidemiol. (Bucur). 16, 429–432.

Le Minor, L., Le Minor, S., 1981. [Origin and frequency of the serotypes of Salmonella isolated in France and received in the French National Center during the years 1977-1979 (author's transl)]. Rev. Epidemiol. Sante Publique 29, 45–55.

Lee, H.-Y., Yang, Y.-J., Su, L.-H., Hsu, C.-H., Fu, Y.-M., Chiu, C.-H., 2008. Genotyping and antimicrobial susceptibility of Salmonella enterica serotype Panama isolated in Taiwan. J. Microbiol. Immunol. Infect. 41, 507–512.

Lee, J.A., 1974. Recent trends in human salmonellosis in England and Wales: the epidemiology of prevalent serotypes other than Salmonella typhimurium. J. Hyg. (Lond). 72, 185–195.

Leeder, F.S., 1956. An epidemic of Salmonella panama infections in infants. Ann. N. Y. Acad. Sci. 66, 54–60.

Lees, J., 2017. The background of bacterial GWAS. https://doi.org/10.6084/m9.figshare.5550037.v1

Lees, J.A., Bentley, S.D., 2016. Bacterial GWAS: not just gilding the lily. Nat. Rev. Microbiol. 14, 406. https://doi.org/10.1038/nrmicro.2016.82

Lees, J.A., Galardini, M., Bentley, S.D., Weiser, J.N., Corander, J., 2018. pyseer: a comprehensive tool for microbial pangenome-wide association studies. Bioinformatics 34, 4310–4312. https://doi.org/10.1093/bioinformatics/bty539

Lees, J.A., Tien Mai, T., Galardini, M., Wheeler, N.E., Corander, J., 2019. Improved inference and prediction of bacterial genotype-phenotype associations using pangenome-spanning regressions. bioRxiv 852426. https://doi.org/10.1101/852426

Leinonen, R., Sugawara, H., Shumway, M., 2011. The sequence read archive. Nucleic Acids Res. 39, D19-21. https://doi.org/10.1093/nar/gkq1019

Leonard, S.R., Lacher, D.W., Lampel, K.A., 2015. Acquisition of the lac operon by Salmonella enterica. BMC Microbiol. 15, 173. https://doi.org/10.1186/s12866-015-0511-8

Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242-5. https://doi.org/10.1093/nar/gkw290

Lewis, J.A., Escalante-Semerena, J.C., 2006. The FAD-dependent tricarballylate dehydrogenase (TcuA) enzyme of Salmonella enterica converts tricarballylate into cis-aconitate. J. Bacteriol. 188, 5479–5486. https://doi.org/10.1128/JB.00514-06

Lewis, J.A., Horswill, A.R., Schwem, B.E., Escalante-Semerena, J.C., 2004. The Tricarballylate utilization (tcuRABC) genes of Salmonella enterica serovar Typhimurium LT2. J. Bacteriol. 186, 1629–1637.

Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

https://doi.org/10.1093/bioinformatics/btp352

Li, R., Yu, C., Li, Y., Lam, T.-W., Yiu, S.-M., Kristiansen, K., Wang, J., 2009. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25, 1966–1967. https://doi.org/10.1093/bioinformatics/btp336

Li, X., Nie, C., Liu, Y., Chen, Y., Lv, X., Wang, L., Zhang, J., Li, K., Jia, Y., Ban, L., Ning, Z., Qu, L., 2019. A genome-wide association study explores the genetic determinism of host resistance to Salmonella pullorum infection in chickens. Genet. Sel. Evol. 51, 51. https://doi.org/10.1186/s12711-019-0492-4

Liao, J., Orsi, R.H., Carroll, L.M., Wiedmann, M., 2020. Comparative genomics reveals different population structures associated with host and geographic origin in antimicrobial-resistant Salmonella enterica. Environ. Microbiol. 22, 2811–2828. https://doi.org/10.1111/1462-2920.15014

Lindstedt, B.-A., Vardund, T., Aas, L., Kapperud, G., 2004. Multiple-locus variable-number tandem-repeats analysis of Salmonella enterica subsp. enterica serovar Typhimurium using PCR multiplexing and multicolor capillary electrophoresis. J. Microbiol. Methods 59, 163–172. https://doi.org/10.1016/j.mimet.2004.06.014

Liu, W.-Q., Feng, Y., Wang, Y., Zou, Q.-H., Chen, F., Guo, J.-T., Peng, Y.-H., Jin, Y., Li, Y.-G., Hu, S.-N., Johnston, R.N., Liu, G.-R., Liu, S.-L., 2009. Salmonella paratyphi C: Genetic Divergence from Salmonella choleraesuis and Pathogenic Convergence with Salmonella typhi. PLoS One 4, e4510.

Lokken, K.L., Walker, G.T., Tsolis, R.M., 2016. Disseminated infections with antibiotic-resistant non-typhoidal Salmonella strains: contributions of host and pathogen factors. Pathog. Dis. 74, 103. https://doi.org/10.1093/femspd/ftw103

Loman, N.J., Pallen, M.J., 2015. Twenty years of bacterial genome sequencing. Nat. Rev. Microbiol. 13, 787—794. https://doi.org/10.1038/nrmicro3565

Lopez-Garrido, J., Cheng, N., Garcia-Quintanilla, F., Garcia-del Portillo, F., Casadesus, J., 2010. Identification of the Salmonella enterica damX gene product, an inner membrane protein involved in bile resistance. J. Bacteriol. 192, 893–895. https://doi.org/10.1128/JB.01220-09

Lopez, C.A., Rivera-Chavez, F., Byndloss, M.X., Baumler, A.J., 2015. The Periplasmic Nitrate Reductase NapABC Supports Luminal Growth of Salmonella enterica Serovar Typhimurium during Colitis. Infect. Immun. 83, 3470–3478. https://doi.org/10.1128/IAI.00351-15

Lukac, M., Pedersen, K., Prukner-Radovcic, E., 2015. Prevalence of Salmonella in captive reptiles from Croatia. J. zoo Wildl. Med. 46, 234–240. https://doi.org/10.1638/2014-0098R1.1

Ma, M., Eaton, J.W., 1992. Multicellular oxidant defense in unicellular organisms. Proc. Natl. Acad. Sci. U. S. A. 89, 7924–7928. https://doi.org/10.1073/pnas.89.17.7924

MacKenzie, K.D., Wang, Y., Musicha, P., Hansen, E.G., Palmer, M.B., Herman, D.J., Feasey, N.A., White, A.P., 2019. Parallel evolution leading to impaired biofilm formation in invasive Salmonella strains. PLOS Genet. 15, e1008233. https://doi.org/10.1371/journal.pgen.1008233

Magiorakos, A.-P., Srinivasan, A., Carey, R.B., Carmeli, Y., Falagas, M.E., Giske, C.G., Harbarth, S., Hindler, J.F., Kahlmeter, G., Olsson-Liljequist, B., Paterson, D.L., Rice, L.B., Stelling, J., Struelens, M.J., Vatopoulos, A., Weber, J.T., Monnet, D.L., 2012. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clin. Microbiol. Infect. 18, 268–281. https://doi.org/10.1111/j.1469-0691.2011.03570.x

Maiden, M.C.J., Jansen van Rensburg, M.J., Bray, J.E., Earle, S.G., Ford, S.A., Jolley, K.A., McCarthy, N.D., 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat. Rev. Microbiol. https://doi.org/10.1038/nrmicro3093

Majowicz, S.E., Musto, J., Scallan, E., Angulo, F.J., Kirk, M., O'Brien, S.J., Jones, T.F., Fazil, A., Hoekstra, R.M., 2010. The global burden of nontyphoidal Salmonella gastroenteritis. Clin. Infect. Dis. 50, 882–889. https://doi.org/10.1086/650733

Malorny, B., Paccassoni, E., Fach, P., Bunge, C., Martin, A., Helmuth, R., 2004. Diagnostic real-time PCR for detection of Salmonella in food. Appl. Environ. Microbiol. 70, 7046–7052. https://doi.org/10.1128/AEM.70.12.7046-7052.2004

Manten, A., Guinee, P.A., Kampelmacher, E.H., Voogd, C.E., 1971. An eleven-year study of drug resistance in Salmonella in the Netherlands. Bull. World Health Organ. 45, 85–93.

Marchello, C.S., Dale, A.P., Pisharody, S., Rubach, M.P., Crump, J.A., 2019. A Systematic Review and Meta-analysis of the Prevalence of Community-Onset Bloodstream Infections among Hospitalized Patients in Africa and Asia. Antimicrob. Agents Chemother. 64, e01974-19. https://doi.org/10.1128/AAC.01974-19

Mastroeni, P., 2002. Immunity to systemic Salmonella infections. Curr. Mol. Med. 2, 393–406. https://doi.org/10.2174/1566524023362492

Mather, A.E., Lawson, B., de Pinna, E., Wigley, P., Parkhill, J., Thomson, N.R., Page, A.J., Holmes, M.A., Paterson, G.K., 2016. Genomic Analysis of Salmonella enterica Serovar Typhimurium from Wild Passerines in England and Wales. Appl. Environ. Microbiol. 82, 6728–6735. https://doi.org/10.1128/AEM.01660-16

Mather, A.E., Matthews, L., Mellor, D.J., Reeve, R., Denwood, M.J., Boerlin, P., Reid-Smith, R.J., Brown, D.J., Coia, J.E., Browning, L.M., Haydon, D.T., Reid, S.W.J., 2012. An ecological approach to assessing the epidemiology of antimicrobial resistance in animal and human populations. Proc. R. Soc. B Biol. Sci. 279, 1630–1639. https://doi.org/10.1098/rspb.2011.1975

Mather, A.E., Phuong, T.L.T., Gao, Y., Clare, S., Mukhopadhyay, S., Goulding, D.A., Hoang, N.T. Do, Tuyen, H.T., Lan, N.P.H., Thompson, C.N., Trang, N.H.T., Carrique-Mas, J., Tue, N.T., Campbell, J.I., Rabaa, M.A., Thanh, D.P., Harcourt, K., Hoa, N.T., Trung, N.V., Schultsz, C., Perron, G.G., Coia, J.E., Brown, D.J., Okoro, C., Parkhill, J., Thomson, N.R., Chau, N.V.V., Thwaites, G.E., Maskell, D.J., Dougan, G., Kenney, L.J., Baker, S., 2018. New Variant of Multidrug-Resistant Serovar Typhimurium Associated with Invasive Disease in Immunocompromised Patients in Vietnam. MBio 9, e01056-18. https://doi.org/10.1128/mBio.01056-18

Mather, A.E., Reid, S.W.J., Maskell, D.J., Parkhill, J., Fookes, M.C., Harris, S.R., Brown, D.J., Coia, J.E., Mulvey, M.R., Gilmour, M.W., Petrovska, L., de Pinna, E., Kuroda, M., Akiba, M., Izumiya, H., Connor, T.R., Suchard, M.A., Lemey, P., Mellor, D.J., Haydon, D.T., Thomson, N.R., 2013. Distinguishable epidemics of multidrug-resistant Salmonella Typhimurium DT104 in different hosts. Science 341, 1514–1517. https://doi.org/10.1126/science.1240578

Mather, A.E., Vaughan, T.G., French, N.P., 2015. Molecular Approaches to Understanding Transmission and Source Attribution in Nontyphoidal Salmonella and Their Application in Africa. Clin. Infect. Dis. 61 Suppl 4, S259-65. https://doi.org/10.1093/cid/civ727

Matias, C.A.R., Pereira, I.A., de Araújo, M. dos S., Santos, A.F.M., Lopes, R.P., Christakis, S., Rodrigues, D.D.P., Siciliano, S., 2016. Characteristics of Salmonella spp. Isolated from Wild Birds Confiscated in Illegal Trade Markets, Rio de Janeiro, Brazil. Biomed Res. Int. 2016, 1–7. https://doi.org/10.1155/2016/3416864

Matiasovicova, J., Havlickova, H., Sisak, F., Pilousova, L., Rychlik, I., 2011. allB, allantoin utilisation and Salmonella enterica serovar Enteritidis and Typhimurium colonisation of poultry and mice. Folia Microbiol. (Praha). 56, 264–269. https://doi.org/10.1007/s12223-011-0034-0

Matsushita, S., Yamada, S., Inaba, M., Kusunoki, J., Kudoh, Y., Ohashi, M., Kawamura, M., Takahashi, M., Yokoyama, K., Konishi, N., Yanagawa, Y., Kai, A., Yamada, S., Morozumi, S., Kudoh, Y., 2001. [Serovar-distribution and drug-resistance of Salmonella strains isolated from domestic and imported cases during 1995-1999 in Tokyo]. Kansenshogaku Zasshi. 75, 116–123.

Matuschek, E., Brown, D.F.J., Kahlmeter, G., 2014. Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. Clin. Microbiol. Infect. 20, O255-66. https://doi.org/10.1111/1469-0691.12373

Maury, M.M., Tsai, Y.-H., Charlier, C., Touchon, M., Chenal-Francisque, V., Leclercq, A., Criscuolo, A., Gaultier, C., Roussel, S., Brisabois, A., Disson, O., Rocha, E.P.C., Brisse, S., Lecuit, M., 2016. Uncovering Listeria monocytogenes hypervirulence by harnessing its biodiversity. Nat. Genet. 48, 308–313. https://doi.org/10.1038/ng.3501

McClelland, M., Sanderson, K.E., Clifton, S.W., Latreille, P., Porwollik, S., Sabo, A., Meyer, R., Bieri, T., Ozersky, P., McLellan, M., Harkins, C.R., Wang, C., Nguyen, C., Berghoff, A., Elliott, G., Kohlberg, S., Strong, C., Du, F., Carter, J., Kremizki, C., Layman, D., Leonard, S., Sun, H., Fulton, L., Nash, W., Miner, T., Minx, P., Delehaunty, K., Fronick, C., Magrini, V., Nhan, M., Warren, W., Florea, L., Spieth, J., Wilson, R.K., 2004.

Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid. Nat. Genet. 36, 1268–1274. https://doi.org/10.1038/ng1470

McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R., Wilson, R.K., 2001. Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. Nature 413, 852–856. https://doi.org/10.1038/35101614

McDonough, P.L., Shin, S.J., Lein, D.H., 2000. Diagnostic and public health dilemma of lactose-fermenting Salmonella enterica serotype Typhimurium in cattle in the Northeastern United States. J. Clin. Microbiol. 38, 1221–1226.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110

McNally, A., Thomson, N.R., Reuter, S., Wren, B.W., 2016. "Add, stir and reduce": Yersinia spp. as model bacteria for pathogen evolution. Nat. Rev. Microbiol. 14, 177—190. https://doi.org/10.1038/nrmicro.2015.29

McQuiston, J.R., Parrenas, R., Ortiz-Rivera, M., Gheesling, L., Brenner, F., Fields, P.I., 2004. Sequencing and comparative analysis of flagellin genes fliC, fljB, and flpA from Salmonella. J. Clin. Microbiol. 42, 1923–1932. https://doi.org/10.1128/jcm.42.5.1923-1932.2004

Mellor, K.C., Petrovska, L., Thomson, N.R., Harris, K., Reid, S.W.J., Mather, A.E., 2019. Antimicrobial Resistance Diversity Suggestive of Distinct Salmonella Typhimurium Sources or Selective Pressures in Food-Production Animals. Front. Microbiol. 10, 708. https://doi.org/10.3389/fmicb.2019.00708

Mensah, N., Tang, Y., Cawthraw, S., AbuOun, M., Fenner, J., Thomson, N.R., Mather, A.E., Petrovska-Holmes, L., 2019. Determining antimicrobial susceptibility in Salmonella enterica serovar Typhimurium through whole genome sequencing: a comparison against multiple phenotypic susceptibility testing methods. BMC Microbiol. 19, 148. https://doi.org/10.1186/s12866-019-1520-9

Merhej, V., Georgiades, K., Raoult, D., 2013. Postgenomic analysis of bacterial pathogens repertoire reveals genome reduction rather than virulence factors. Brief. Funct. Genomics 12, 291—304. https://doi.org/10.1093/bfgp/elt015

Mermin, J., Hutwagner, L., Vugia, D., Shallow, S., Daily, P., Bender, J., Koehler, J., Marcus, R., Angulo, F.J., 2004. Reptiles, amphibians, and human Salmonella infection: a population-based, case-control study. Clin. Infect. Dis. 38 Suppl 3, S253-61. https://doi.org/10.1086/381594

Miller, J.R., Koren, S., Sutton, G., 2010. Assembly algorithms for next-generation sequencing data. Genomics 95, 315–327. https://doi.org/https://doi.org/10.1016/j.ygeno.2010.03.001

Modai, J., Robineau, M., Brucker, G., Veyssier, P., Neveux, J.Y., Domart, A., 1974. [Septicemia caused by Salmonella panama and aneurysm of the descending aorta]. Ann. Med. Interne (Paris). 125, 581–585.

Mogasale, Vittal, Maskery, B., Ochiai, R.L., Lee, J.S., Mogasale, Vijayalaxmi V, Ramani, E., Kim, Y.E., Park, J.K., Wierzba, T.F., 2014. Burden of typhoid fever in low-income and middle-income countries: a systematic, literature-based update with risk-factor adjustment. Lancet. Glob. Heal. 2, e570-80. https://doi.org/10.1016/S2214-109X(14)70301-8

Monteiro, C., Fang, X., Ahmad, I., Gomelsky, M., Römling, U., 2011. Regulation of biofilm components in Salmonella enterica serovar Typhimurium by lytic transglycosylases involved in cell wall turnover. J. Bacteriol. 193, 6443–6451. https://doi.org/10.1128/JB.00425-11

Moreira, C.G., Herrera, C.M., Needham, B.D., Parker, C.T., Libby, S.J., Fang, F.C., Trent, M.S., Sperandio, V., 2013. Virulence and stress-related periplasmic protein (VisP) in bacterial/host associations. Proc. Natl. Acad. Sci. 110, 1470 LP – 1475. https://doi.org/10.1073/pnas.1215416110

Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T.L., Agarwala, R., Schaffer, A.A., 2008. Database indexing for production MegaBLAST searches. Bioinformatics 24, 1757–1764.

https://doi.org/10.1093/bioinformatics/btn322

Mottawea, W., Duceppe, M.-O., Dupras, A.A., Usongo, V., Jeukens, J., Freschi, L., Emond-Rheault, J.-G., Hamel, J., Kukavica-Ibrulj, I., Boyle, B., Gill, A., Burnett, E., Franz, E., Arya, G., Weadge, J.T., Gruenheid, S., Wiedmann, M., Huang, H., Daigle, F., Moineau, S., Bekal, S., Levesque, R.C., Goodridge, L.D., Ogunremi, D., 2018. Salmonella enterica Prophage Sequence Profiles Reflect Genome Diversity and Can Be Used for High Discrimination Subtyping. Front. Microbiol. 9, 836. https://doi.org/10.3389/fmicb.2018.00836

Moustafa, A.M., Lal, A., Planet, P.J., 2020. Comparative genomics in infectious disease. Curr. Opin. Microbiol. 53, 61–70. https://doi.org/10.1016/j.mib.2020.02.009

Müller, N.F., Rasmussen, D., Stadler, T., 2018. MASCOT: parameter and state inference under the marginal structured coalescent approximation. Bioinformatics 34, 3843–3848. https://doi.org/10.1093/bioinformatics/bty406

Murphy, D., Oshin, F., 2015. Reptile-associated salmonellosis in children aged under 5 years in South West England. Arch. Dis. Child. 100, 364–365. https://doi.org/10.1136/archdischild-2014-306134

Nair, S., Wain, J., Connell, S., de Pinna, E., Peters, T., 2014. Salmonella enterica subspecies II infections in England and Wales--the use of multilocus sequence typing to assist serovar identification. J. Med. Microbiol. 63, 831–834. https://doi.org/10.1099/jmm.0.072702-0

Nakadai, A., Kuroki, T., Kato, Y., Suzuki, R., Yamai, S., Yaginuma, C., Shiotani, R., Yamanouchi, A., Hayashidani, H., 2005. Prevalence of Salmonella spp. in pet reptiles in Japan. J. Vet. Med. Sci. 67, 97–101.

Neuert, S., Nair, S., Day, M.R., Doumith, M., Ashton, P.M., Mellor, K.C., Jenkins, C., Hopkins, K.L., Woodford, N., de Pinna, E., Godbole, G., Dallman, T.J., 2018. Prediction of Phenotypic Antimicrobial Resistance Profiles From Whole Genome Sequences of Non-typhoidal Salmonella enterica. Front. Microbiol. 9, 592. https://doi.org/10.3389/fmicb.2018.00592

Nguyen, B.D., Cuenca V., M., Hartl, J., Gül, E., Bauer, R., Meile, S., Rüthi, J., Margot, C., Heeb, L., Besser, F., Escriva, P.P., Fetz, C., Furter, M., Laganenka, L., Keller, P., Fuchs, L., Christen, M., Porwollik, S., McClelland, M., Vorholt, J.A., Sauer, U., Sunagawa, S., Christen, B., Hardt, W.-D., 2020. Import of Aspartate and Malate by DcuABC Drives H2/Fumarate Respiration to Promote Initial Salmonella Gut-Lumen Colonization in Mice. Cell Host Microbe 27, 922-936.e6. https://doi.org/10.1016/j.chom.2020.04.013

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300

Nguyen, K. Van, Thi Do, N.T., Chandna, A., Nguyen, T.V., Pham, C. Van, Doan, P.M., Nguyen, A.Q., Thi Nguyen, C.K., Larsson, M., Escalante, S., Olowokure, B., Laxminarayan, R., Gelband, H., Horby, P., Thi Ngo, H.B., Hoang, M.T., Farrar, J., Hien, T.T., Wertheim, H.F.L., 2013. Antibiotic use and resistance in emerging economies: a situation analysis for Viet Nam. BMC Public Health 13, 1158. https://doi.org/10.1186/1471-2458-13-1158

Nhung, N., Cuong, N., Thwaites, G., Carrique-Mas, J., 2016. Antimicrobial Usage and Antimicrobial Resistance in Animal Production in Southeast Asia: A Review. Antibiotics 5, 37. https://doi.org/10.3390/antibiotics5040037

Nichols, R.J., Sen, S., Choo, Y.J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K.M., Lee, K.J., Wong, A., Shales, M., Lovett, S., Winkler, M.E., Krogan, N.J., Typas, A., Gross, C.A., 2011. Phenotypic landscape of a bacterial cell. Cell 144, 143–156. https://doi.org/10.1016/j.cell.2010.11.052

Nightingale, K.K., Windham, K., Wiedmann, M., 2005. Evolution and Molecular Phylogeny of Listeria monocytogenes Isolated from Human and Animal Listeriosis Cases and Foods. J. Bacteriol. 187, 5537 LP – 5551. https://doi.org/10.1128/JB.187.16.5537-5551.2005

Nishino, K., Latifi, T., Groisman, E.A., 2006. Virulence and drug resistance roles of multidrug efflux systems of Salmonella enterica serovar Typhimurium. Mol. Microbiol. 59, 126–141. https://doi.org/doi:10.1111/j.1365-2958.2005.04940.x

Noël, H., Hofhuis, A., De Jonge, R., Heuvelink, A.E., De Jong, A., Heck, M.E.O.C., De Jager, C., van Pelt, W., 2010.

Consumption of Fresh Fruit Juice: How a Healthy Food Practice Caused a National Outbreak of Salmonella Panama Gastroenteritis. Foodborne Pathog. Dis. 7, 375–381. https://doi.org/10.1089/fpd.2009.0330

Nolle, N., Felsl, A., Heermann, R., Fuchs, T.M., 2017. Genetic Characterization of the Galactitol Utilization Pathway of Salmonella enterica Serovar Typhimurium. J. Bacteriol. 199, e00595-16. https://doi.org/10.1128/JB.00595-16

Nuccio, S.-P., Bäumler, A.J., 2014. Comparative Analysis of *Salmonella* Genomes Identifies a Metabolic Network for Escalating Growth in the Inflamed Gut. MBio 5, e00929-14. https://doi.org/10.1128/mBio.00929-14

Nyrén, P., Lundin, A., 1985. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. Anal. Biochem. 151, 504–509. https://doi.org/10.1016/0003-2697(85)90211-8

O'Neill, J., 2019. Tackling drug-resistant infections globally: final report and recommendations: the review on antimicrobial resistance; 2016 [Available from: https://amr-review. org. Publ. html.

Ogunfowokan, O., 2012. Bite-to-hospital time and morbidity in victims of viper bite in a rural hospital in Nigeria. African J. Prim. Heal. Care Fam. Med. 4, 371. https://doi.org/10.4102/phcfm.v4i1.371

Okoro, C.K., Barquist, L., Connor, T.R., Harris, S.R., Clare, S., Stevens, M.P., Arends, M.J., Hale, C., Kane, L., Pickard, D.J., Hill, J., Harcourt, K., Parkhill, J., Dougan, G., Kingsley, R.A., 2015. Signatures of adaptation in human invasive Salmonella Typhimurium ST313 populations from sub-Saharan Africa. PLoS Negl. Trop. Dis. 9, e0003611. https://doi.org/10.1371/journal.pntd.0003611

Okoro, C.K., Kingsley, R.A., Connor, T.R., Harris, S.R., Parry, C.M., Al-Mashhadani, M.N., Kariuki, S., Msefula, C.L., Gordon, M.A., de Pinna, E., Wain, J., Heyderman, R.S., Obaro, S., Alonso, P.L., Mandomando, I., MacLennan, C.A., Tapia, M.D., Levine, M.M., Tennant, S.M., Parkhill, J., Dougan, G., 2012. Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa. Nat. Genet. 44, 1215–1221. https://doi.org/10.1038/ng.2423

Olive, C., Mansuy, J.M.M., Desbois, N., Roche, B., Cecile, W., Saint-Aime, C., Jouannelle, J., 1996. Salmonella panama en Martinique : aspects épidémiologiques et cliniques chez l'enfant hospitalisé. Médecine Mal. Infect. 26, 590–593. https://doi.org/https://doi.org/10.1016/S0399-077X(96)80078-3

Oneko, M., Kariuki, S., Muturi-Kioi, V., Otieno, K., Otieno, V.O., Williamson, J.M., Folster, J., Parsons, M.B., Slutsker, L., Mahon, B.E., Hamel, M.J., 2015. Emergence of Community-Acquired, Multidrug-Resistant Invasive Nontyphoidal Salmonella Disease in Rural Western Kenya, 2009–2013. Clin. Infect. Dis. 61, S310–S316. https://doi.org/10.1093/cid/civ674

Opree, W., 1975. [Infection with "enteritis salmonella" at non-intestinal sites (author's transl)]. Dtsch. Med. Wochenschr. 100, 1425–1428.

Otsuka, Y., Yonesaki, T., 2012. Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins. Mol. Microbiol. 83, 669–681. https://doi.org/10.1111/j.1365-2958.2012.07975.x

Owen, S. V, Wenner, N., Canals, R., Makumi, A., Hammarlof, D.L., Gordon, M.A., Aertsen, A., Feasey, N.A., Hinton, J.C.D., 2017. Characterization of the Prophage Repertoire of African Salmonella Typhimurium ST313 Reveals High Levels of Spontaneous Induction of Novel Phage BTP1. Front. Microbiol. 8, 235. https://doi.org/10.3389/fmicb.2017.00235

Owen, S. V, Wenner, N., Dulberger, C.L., Rodwell, E. V, Bowers-Barnard, A., Quinones-Olvera, N., Rigden, D.J., Rubin, E.J., Garner, E.C., Baym, M., Hinton, J.C.D., 2020. Prophage-encoded phage defence proteins with cognate self-immunity. bioRxiv 2020.07.13.199331. https://doi.org/10.1101/2020.07.13.199331

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31, 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A., Harris, S.R., 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb. genomics 2, e000056. https://doi.org/10.1099/mgen.0.000056

Pallen, M.J., Wren, B.W., 2007. Bacterial pathogenomics. Nature 449, 835—842.

https://doi.org/10.1038/nature06248

Papa, F., 1976. [Contribution to the study of Salmonella in Martinique. Evolution during 1972, 1973 and 1974]. Bull. Soc. Pathol. Exot. Filiales 69, 121–5.

Parkhill, J., 2013. What has high-throughput sequencing ever done for us? Nat. Rev. Microbiol. 11, 664–665. https://doi.org/10.1038/nrmicro3112

Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T.G., Sebaihia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T., Connerton, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N., Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S., Jagels, K., Krogh, A., Larsen, T.S., Leather, S., Moule, S., Ó'Gaora, P., Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., Barrell, B.G., 2001. Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. Nature 413, 848–852. https://doi.org/10.1038/35101607

Parry, C.M., Ribeiro, I., Walia, K., Rupali, P., Baker, S., Basnyat, B., 2019. Multidrug resistant enteric fever in South Asia: unmet medical needs and opportunities. BMJ 364, k5322–k5322. https://doi.org/10.1136/bmj.k5322

Parsons, S.K., Bull, C.M., Gordon, D.M., 2015. Spatial Variation and Survival of Salmonella enterica Subspecies in a Population of Australian Sleepy Lizards (Tiliqua rugosa). Appl. Environ. Microbiol. 81, 5804–5811. https://doi.org/10.1128/AEM.00997-15

Parsons, S.K., Bull, C.M., Gordon, D.M., 2011. Substructure within Salmonella enterica subsp. enterica isolates from Australian wildlife. Appl. Environ. Microbiol. 77, 3151–3153. https://doi.org/10.1128/AEM.02764-10

Pedersen, K., Lassen-Nielsen, A.-M., Nordentoft, S., Hammer, A.S., 2009. Serovars of Salmonella from captive reptiles. Zoonoses Public Health 56, 238–242. https://doi.org/10.1111/j.1863-2378.2008.01196.x

Pelludat, C., Mirold, S., Hardt, W.-D., 2003. The SopEPhi phage integrates into the ssrA gene of Salmonella enterica serovar Typhimurium A36 and is closely related to the Fels-2 prophage. J. Bacteriol. 185, 5182–5191. https://doi.org/10.1128/jb.185.17.5182-5191.2003

Penner, J.L., 1988. International Committee on Systematic Bacteriology Taxonomic Subcommittee on Enterobacteriaceae: Minutes of the Meeting, 8 September 1986, Manchester, England. Int. J. Syst. Bacteriol. 38, 223–224. https://doi.org/10.1099/00207713-38-2-223

Pérez-Losada, M., Browne, E.B., Madsen, A., Wirth, T., Viscidi, R.P., Crandall, K.A., 2006. Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 6, 97–112. https://doi.org/10.1016/j.meegid.2005.02.003

Perez-Sepulveda, B.M., Heavens, D., Pulford, C. V, Predeus, A. V, Low, R., Webster, H., Schudoma, C., Rowe, W., Lipscombe, J., Watkins, C., Kumwenda, B., Shearer, N., Costigan, K., Baker, K.S., Feasey, N.A., Hinton, J.C.D., Hall, N., 2020. An accessible, efficient and global approach for the large-scale sequencing of bacterial genomes. bioRxiv 2020.07.22.200840. https://doi.org/10.1101/2020.07.22.200840

Phoba, M.-F., Barbé, B., Ley, B., Van Puyvelde, S., Post, A., Mattheus, W., Deborggraeve, S., Lunguya, O., Jacobs, J., 2020. High genetic similarity between non-typhoidal Salmonella isolated from paired blood and stool samples of children in the Democratic Republic of the Congo. PLoS Negl. Trop. Dis. 14, e0008377–e0008377. https://doi.org/10.1371/journal.pntd.0008377

Phu Huong Lan, N., Le Thi Phuong, T., Nguyen Huu, H., Thuy, L., Mather, A.E., Park, S.E., Marks, F., Thwaites, G.E., Van Vinh Chau, N., Thompson, C.N., Baker, S., 2016. Invasive Non-typhoidal Salmonella Infections in Asia: Clinical Observations, Disease Outcome and Dominant Serovars from an Infectious Disease Hospital in Vietnam. PLoS Negl. Trop. Dis. 10, e0004857.

Piasecki, T., Chrząstek, K., Wieliczko, A., 2014. Salmonella serovar spectrum associated with reptiles in Poland. Acta Vet. Brno 83, 287–294.

Piccini, G., Montomoli, E., 2020. Pathogenic signature of invasive non-typhoidal Salmonella in Africa: implications for vaccine development. Hum. Vaccin. Immunother. 1–16. https://doi.org/10.1080/21645515.2020.1785791

Pijnacker, R., Dallman, T.J., Tijsma, A.S.L., Hawkins, G., Larkin, L., Kotila, S.M., Amore, G., Amato, E., Suzuki, P.M., Denayer, S., Klamer, S., Pászti, J., McCormick, J., Hartman, H., Hughes, G.J., Brandal, L.C.T., Brown, D., Mossong, J., Jernberg, C., Müller, L., Palm, D., Severi, E., Gołębiowska, J., Hunjak, B., Owczarek, S., Le Hello, S., Garvey, P., Mooijman, K., Friesema, I.H.M., van der Weijden, C., van der Voort, M., Rizzi, V., Franz, E., Bertrand, S., Brennan, M., Browning, L., Bruce, R., Cantaert, V., Chattaway, M., Coia, J., Couper, S., Žohar Čretnik, T., Daniel, O., Dionisi, A.M., Fabre, L., Fitz-James, I., Florek, K., Florianová, M., Fox, E., Frelih, T., Grilc, E., Katalinic Jankovic, V., Jourdan, N., Karpíšková, R., van den Kerkhof, H., Kuiling, S., Kurečić Filipović, S., Laisnez, V., Lange, H., DeLappes, N., Leblanc, J., Luzzi, I., Mandilara, G., Mather, H., Mattheus, W., Mellou, K., Morgan, D., Pászti, J., de Pinna, E., Ragimbeau, C., Røed, M.H., Salmenlinna, S., Smith, R., Smith-Palmer, A., Špačková, M., Torpdahl, M., Trkov, M., Trönnberg, L., Tzani, M., Utsi, L., Wasyl, D., Weicherding, P., 2019. An international outbreak of Salmonella enterica serotype Enteritidis linked to eggs from Poland: a microbiological and epidemiological study. Lancet Infect. Dis. 19, 778–786. https://doi.org/10.1016/S1473-3099(19)30047-7

Porwollik, S., Santiviago, C.A., Cheng, P., Florea, L., McClelland, M., 2005. Differences in Gene Content between Salmonella enterica Serovar Enteritidis Isolates and Comparison to Closely Related Serovars Gallinarum and Dublin. J. Bacteriol. 187, 6545 LP – 6555. https://doi.org/10.1128/JB.187.18.6545-6555.2005

Post, A.S., Diallo, S.N., Guiraud, I., Lompo, P., Tahita, M.C., Maltha, J., Van Puyvelde, S., Mattheus, W., Ley, B., Thriemer, K., Rouamba, E., Derra, K., Deborggraeve, S., Tinto, H., Jacobs, J., 2019. Supporting evidence for a human reservoir of invasive non-Typhoidal Salmonella from household samples in Burkina Faso. PLoS Negl. Trop. Dis. 13, e0007782.

Power, R.A., Parkhill, J., de Oliveira, T., 2017. Microbial genome-wide association studies: lessons from human GWAS. Nat. Rev. Genet. 18, 41–50. https://doi.org/10.1038/nrg.2016.132

Public Health England, 2014. Detection of Salmonella species. Microbiology Services. Food, Water & Environmental Microbiology Standard Method FNES16 (F13). Version 2. Public Health England, London.

Pulford, C. V, Perez-Sepulveda, B.M., Rodwell, E. V, Weill, F.-X., Baker, K.S., Hinton, J.C.D., 2019. Salmonella enterica Serovar Panama, an Understudied Serovar Responsible for Extraintestinal Salmonellosis Worldwide. Infect. Immun. 87, e00273-19. https://doi.org/10.1128/IAI.00273-19

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., Gu, Y., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13, 341. https://doi.org/10.1186/1471-2164-13-341

R Core Team, 2017. R: A language and environment for statistical computing.

Rabsch, W., 2007. Salmonella typhimurium phage typing for pathogens. Methods Mol. Biol. 394, 177–211. https://doi.org/10.1007/978-1-59745-512-1_10

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst. Biol. 67, 901–904. https://doi.org/10.1093/sysbio/syy032

Rambaut, A., Lam, T.T., Max Carvalho, L., Pybus, O.G., 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2, vew007–vew007. https://doi.org/10.1093/ve/vew007

Reddy, E.A., Shaw, A. V., Crump, J.A., 2010. Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis. Lancet. Infect. Dis. 10, 417–432. https://doi.org/10.1016/S1473-3099(10)70072-4

Reuter, S., Connor, T.R., Barquist, L., Walker, D., Feltwell, T., Harris, S.R., Fookes, M., Hall, M.E., Petty, N.K., Fuchs, T.M., Corander, J., Dufour, M., Ringwood, T., Savin, C., Bouchier, C., Martin, L., Miettinen, M., Shubin, M., Riehm, J.M., Laukkanen-Ninios, R., Sihvonen, L.M., Siitonen, A., Skurnik, M., Falcão, J.P., Fukushima, H., Scholz, H.C., Prentice, M.B., Wren, B.W., Parkhill, J., Carniel, E., Achtman, M., McNally, A., Thomson, N.R., 2014. Parallel independent evolution of pathogenicity within the genus Yersinia. Proc. Natl. Acad. Sci. U. S. A. 111, 6768—6773. https://doi.org/10.1073/pnas.1317161111

Rhoads, A., Au, K.F., 2015. PacBio Sequencing and Its Applications. Genomics. Proteomics Bioinformatics 13, 278–289. https://doi.org/https://doi.org/10.1016/j.gpb.2015.08.002

Ribot, E.M., Freeman, M., Hise, K.B., Gerner-Smidt, P., 2019. PulseNet: Entering the Age of Next-Generation Sequencing. Foodborne Pathog. Dis. 16, 451–456. https://doi.org/10.1089/fpd.2019.2634

Ricke, S.C., Calo, J.R., Kaldhone, P., 2015. Chapter 7 - Salmonella Control in Food Production: Current Issues and Perspectives in the United States, in: Ricke, S.C., Donaldson, J.R., Phillips, C.A.B.T.-F.S. (Eds.), . Academic Press, San Diego, pp. 107–133. https://doi.org/https://doi.org/10.1016/B978-0-12-800245-2.00007-1

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. https://doi.org/10.1038/nbt.1754

Rodríguez, I., Martín, M.C., Mendoza, M.C., Rodicio, M.R., 2006. Class 1 and class 2 integrons in non-prevalent serovars of Salmonella enterica: Structure and association with transposons and plasmids. J. Antimicrob. Chemother. 58, 1124–1132. https://doi.org/10.1093/jac/dkl400

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542. https://doi.org/10.1093/sysbio/sys029

Rotger, R., Casadesus, J., 1999. The virulence plasmids of Salmonella. Int. Microbiol. 2, 177–184.

Ryan, M.P., O'Dwyer, J., Adley, C.C., 2017. Evaluation of the Complex Nomenclature of the Clinically and Veterinary Significant Pathogen *Salmonella*. Biomed Res. Int. 2017, 3782182. https://doi.org/10.1155/2017/3782182

Sá, I.V.A. de, Solari, C.A., 2001. Salmonella in Brazilian and imported pet reptiles. Brazilian J. Microbiol. 32, 293–297. https://doi.org/10.1590/S1517-83822001000400007

Sahin, F., Karasartova, D., Gerceker, D., Aysev, A.D., Erdem, B., 2008. [A novel Salmonella Typhimurium plasmid, pAnkS: an example for plasmid evolution in antibiotic resistance]. Mikrobiyol. Bul. 42, 383–388.

Sahl, J.W., Lemmer, D., Travis, J., Schupp, J.M., Gillece, J.D., Aziz, M., Driebe, E.M., Drees, K.P., Hicks, N.D., Williamson, C.H.D., Hepp, C.M., Smith, D.E., Roe, C., Engelthaler, D.M., Wagner, D.M., Keim, P., 2016. NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. Microb. genomics 2, e000074–e000074. https://doi.org/10.1099/mgen.0.000074

Salamon, S.A., Prag, J., 2001. A case of superficial septic thrombophlebitis in a varicose vein caused by Salmonella panama. Clin. Microbiol. Infect. 7, 34–36.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Schatz, M.C., Delcher, A.L., Salzberg, S.L., 2010. Assembly of large genomes using second-generation sequencing. Genome Res. 20, 1165–1173. https://doi.org/10.1101/gr.101360.109

Scheurwater, E., Reid, C.W., Clarke, A.J., 2008. Lytic transglycosylases: Bacterial space-making autolysins. Int. J. Biochem. Cell Biol. 40, 586–591. https://doi.org/https://doi.org/10.1016/j.biocel.2007.03.018

Schroter, M., Roggentin, P., Hofmann, J., Speicher, A., Laufs, R., Mack, D., 2004. Pet snakes as a reservoir for Salmonella enterica subsp. diarizonae (Serogroup IIIb): a prospective study. Appl. Environ. Microbiol. 70, 613–615.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Seif, Y., Kavvas, E., Lachance, J.-C., Yurkovich, J.T., Nuccio, S.-P., Fang, X., Catoiu, E., Raffatellu, M., Palsson, B.O., Monk, J.M., 2018. Genome-scale metabolic reconstructions of multiple Salmonella strains reveal serovar-specific metabolic traits. Nat. Commun. 9, 3771. https://doi.org/10.1038/s41467-018-06112-5

Selander, R.K., Beltran, P., Smith, N.H., Helmuth, R., Rubin, F.A., Kopecko, D.J., Ferris, K., Tall, B.D., Cravioto, A., Musser, J.M., 1990. Evolutionary genetic relationships of clones of Salmonella serovars that cause human typhoid and other enteric fevers. Infect. Immun. 58, 2262–2275.

Shane, A.L., Mody, R.K., Crump, J.A., Tarr, P.I., Steiner, T.S., Kotloff, K., Langley, J.M., Wanke, C., Warren, C.A., Cheng, A.C., Cantey, J., Pickering, L.K., 2017. 2017 Infectious Diseases Society of America Clinical Practice

Guidelines for the Diagnosis and Management of Infectious Diarrhea. Clin. Infect. Dis. 65, e45–e80. https://doi.org/10.1093/cid/cix669

Shendure, J., Ji, H., 2008. Next-generation DNA sequencing. Nat. Biotechnol. 26, 1135–1145. https://doi.org/10.1038/nbt1486

Shivaprasad, H.L., 2000. Fowl typhoid and pullorum disease. Rev. Sci. Tech. 19, 405–424. https://doi.org/10.20506/rst.19.2.1222

Silva, C., Puente, J.L., Calva, E., 2017. Salmonella virulence plasmid: pathogenesis and ecology. Pathog. Dis. 75. https://doi.org/10.1093/femspd/ftx070

Simpson, K.M.J., Hill-Cawthorne, G.A., Ward, M.P., Mor, S.M., 2018. Diversity of Salmonella serotypes from humans, food, domestic animals and wildlife in New South Wales, Australia. BMC Infect. Dis. 18, 623. https://doi.org/10.1186/s12879-018-3563-1

Singletary, L.A., Karlinsey, J.E., Libby, S.J., Mooney, J.P., Lokken, K.L., Tsolis, R.M., Byndloss, M.X., Hirao, L.A., Gaulke, C.A., Crawford, R.W., Dandekar, S., Kingsley, R.A., Msefula, C.L., Heyderman, R.S., Fang, F.C., 2016. Loss of Multicellular Behavior in Epidemic African Nontyphoidal Salmonella enterica Serovar Typhimurium ST313 Strain D23580. MBio 7, e02265. https://doi.org/10.1128/mBio.02265-15

Soto, S.M., Guerra, B., Del Cerro, A., González-Hevia, M.A., Mendoza, M.C., 2001. Outbreaks and sporadic cases of Salmonella serovar Panama studied by DNA fingerprinting and antimicrobial resistance. Int. J. Food Microbiol. 71, 35–43. https://doi.org/10.1016/S0168-1605(01)00553-0

Srikumar, S., Kröger, C., Hébrard, M., Colgan, A., Owen, S. V, Sivasankaran, S.K., Cameron, A.D.S., Hokamp, K., Hinton, J.C.D., 2015. RNA-seq Brings New Insights to the Intra-Macrophage Transcriptome of Salmonella Typhimurium. PLOS Pathog. 11, e1005262.

Srividhya, K. V, Krishnaswamy, S., 2007. Subclassification and targeted characterization of prophage-encoded two-component cell lysis cassette. J. Biosci. 32, 979–990. https://doi.org/10.1007/s12038-007-0097-x

Staden, R., 1979. A strategy of DNA sequencing employing computer programs. Nucleic Acids Res. 6, 2601–2610. https://doi.org/10.1093/nar/6.7.2601

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stanaway, J.D., Parisi, A., Sarkar, K., Blacker, B.F., Reiner, R.C., Hay, S.I., Nixon, M.R., Dolecek, C., James, S.L., Mokdad, A.H., Abebe, G., Ahmadian, E., Alahdab, F., Alemnew, B.T.T., Alipour, V., Allah Bakeshei, F., Animut, M.D., Ansari, F., Arabloo, J., Asfaw, E.T., Bagherzadeh, M., Bassat, Q., Belayneh, Y.M.M., Carvalho, F., Daryani, A., Demeke, F.M., Demis, A.B.B., Dubey, M., Duken, E.E., Dunachie, S.J., Eftekhari, A., Fernandes, E., Fouladi Fard, R., Gedefaw, G.A., Geta, B., Gibney, K.B., Hasanzadeh, A., Hoang, C.L., Kasaeian, A., Khater, A., Kidanemariam, Z.T., Lakew, A.M., Malekzadeh, R., Melese, A., Mengistu, D.T., Mestrovic, T., Miazgowski, B., Mohammad, K.A., Mohammadian, M., Mohammadian-Hafshejani, A., Nguyen, C.T., Nguyen, L.H., Nguyen, S.H., Nirayo, Y.L., Olagunju, A.T., Olagunju, T.O., Pourjafar, H., Qorbani, M., Rabiee, M., Rabiee, N., Rafay, A., Rezapour, A., Samy, A.M., Sepanlou, S.G., Shaikh, M.A., Sharif, M., Shigematsu, M., Tessema, B., Tran, B.X., Ullah, I., Yimer, E.M., Zaidi, Z., Murray, C.J.L., Crump, J.A., 2019a. The global burden of non-typhoidal salmonella invasive disease: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Infect. Dis. 19, 1312–1324. https://doi.org/10.1016/S1473-3099(19)30418-9

Stanaway, J.D., Reiner, R.C., Blacker, B.F., Goldberg, E.M., Khalil, I.A., Troeger, C.E., Andrews, J.R., Bhutta, Z.A., Crump, J.A., Im, J., Marks, F., Mintz, E., Park, S.E., Zaidi, A.K.M., Abebe, Z., Abejie, A.N., Adedeji, I.A., Ali, B.A., Amare, A.T., Atalay, H.T., Avokpaho, E.F.G.A., Bacha, U., Barac, A., Bedi, N., Berhane, A., Browne, A.J., Chirinos, J.L., Chitheer, A., Dolecek, C., El Sayed Zaki, M., Eshrati, B., Foreman, K.J., Gemechu, A., Gupta, R., Hailu, G.B., Henok, A., Hibstu, D.T., Hoang, C.L., Ilesanmi, O.S., Iyer, V.J., Kahsay, A., Kasaeian, A., Kassa, T.D., Khan, E.A., Khang, Y.-H., Magdy Abd El Razek, H., Melku, M., Mengistu, D.T., Mohammad, K.A., Mohammed, S., Mokdad, A.H., Nachega, J.B., Naheed, A., Nguyen, C.T., Nguyen, H.L.T., Nguyen, L.H., Nguyen, N.B., Nguyen, T.H., Nirayo, Y.L., Pangestu, T., Patton, G.C., Qorbani, M., Rai, R.K., Rana, S.M., Ranabhat, C.L., Roba, K.T., Roberts, N.L.S., Rubino, S., Safiri, S., Sartorius, B., Sawhney, M., Shiferaw, M.S., Smith, D.L., Sykes, B.L., Tran, B.X., Tran, T.T., Ukwaja, K.N., Vu, G.T., Vu, L.G., Weldegebreal, F., Yenit, M.K., Murray, C.J.L., Hay, S.I., 2019b. The global burden of typhoid and

paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Infect. Dis. 19, 369–381. https://doi.org/10.1016/S1473-3099(18)30685-6

Stanley, J., Baquar, N., Burnens, A., 1995. Molecular subtyping scheme for Salmonella panama. J. Clin. Microbiol. 33, 1206–1211.

Starai, V.J., Garrity, J., Escalante-Semerena, J.C., 2005. Acetate excretion during growth of Salmonella enterica on ethanolamine requires phosphotransacetylase (EutD) activity, and acetate recapture requires acetyl-CoA synthetase (Acs) and phosphotransacetylase (Pta) activities. Microbiology 151, 3793–3801. https://doi.org/10.1099/mic.0.28156-0

Steeb, B., Claudi, B., Burton, N.A., Tienz, P., Schmidt, A., Farhan, H., Mazé, A., Bumann, D., 2013. Parallel Exploitation of Diverse Host Nutrients Enhances Salmonella Virulence. PLOS Pathog. 9, e1003301.

Stephan, R., Bulling, E., Steinbeck, A., 1977. [The development of antibiotics resistance among Salmonella bacteria of animal origin in the Federal Republic of Germany and Berlin (West). 6th communication: 1975 Annual report (author's transl)]. Zentralbl. Bakteriol. Orig. A. 237, 264–273.

Stocker, P., Rosner, B., Werber, D., Kirchner, M., Reinecke, A., Wichmann-Schauer, H., Prager, R., Rabsch, W., Frank, C., 2011. Outbreak of Salmonella Montevideo associated with a dietary food supplement flagged in the Rapid Alert System for Food and Feed (RASFF) in Germany, 2010. Euro Surveill. Eur. Commun. Dis. Bull. 16, 20040.

Susskind, M.M., Wright, A., Botstein, D., 1974. Superinfection exclusion by P22 prophage in lysogens of Salmonella typhimurium: IV. Genetics and physiology of sieB exclusion. Virology 62, 367–384. https://doi.org/https://doi.org/10.1016/0042-6822(74)90399-7

Svennblad, B., Erixon, P., Oxelman, B., Britton, T., 2006. Fundamental Differences Between the Methods of Maximum Likelihood and Maximum Posterior Probability in Phylogenetics. Syst. Biol. 55, 116–121. https://doi.org/10.1080/10635150500481648

Swaminathan, B., Barrett, T.J., Hunter, S.B., Tauxe, R. V, Force, C.D.C.P.T., 2001. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg. Infect. Dis. 7, 382–389. https://doi.org/10.3201/eid0703.010303

Tack, B., Vanaenrode, J., Verbakel, J.Y., Toelen, J., Jacobs, J., 2020. Invasive non-typhoidal Salmonella infections in sub-Saharan Africa: a systematic review on antimicrobial resistance and treatment. BMC Med. 18, 212. https://doi.org/10.1186/s12916-020-01652-4

Talon, P., Schneller, E., Stoll, C., 1985. [Salmonella panama responsible for meningitis secondary to febrile gastroenteritis in a 3-month-old infant]. Pediatrie 40, 223–227.

The, H.C., Thanh, D.P., Holt, K.E., Thomson, N.R., Baker, S., 2016. The genomic signatures of Shigella evolution, adaptation and geographical spread. Nat. Rev. Microbiol. 14, 235—250. https://doi.org/10.1038/nrmicro.2016.10

Thomson, N.R., Clayton, D.J., Windhorst, D., Vernikos, G., Davidson, S., Churcher, C., Quail, M.A., Stevens, M., Jones, M.A., Watson, M., Barron, A., Layton, A., Pickard, D., Kingsley, R.A., Bignell, A., Clark, L., Harris, B., Ormond, D., Abdellah, Z., Brooks, K., Cherevach, I., Chillingworth, T., Woodward, J., Norberczak, H., Lord, A., Arrowsmith, C., Jagels, K., Moule, S., Mungall, K., Sanders, M., Whitehead, S., Chabalgoity, J.A., Maskell, D., Humphrey, T., Roberts, M., Barrow, P.A., Dougan, G., Parkhill, J., 2008. Comparative genome analysis of Salmonella Enteritidis PT4 and Salmonella Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. Genome Res. 18, 1624–1637. https://doi.org/10.1101/gr.077404.108

Tindall, B., Grimont, P., Garrity, G., Euzéby, J., 2005. Nomenclature and taxonomy of the genus Salmonella. Int. J. Syst. Evol. Microbiol. 55, 521–524. https://doi.org/10.1099/ijs.0.63580-0

Toba, F.A., Thompson, M.G., Campbell, B.R., Junker, L.M., Rueggeberg, K.-G., Hay, A.G., 2011. Role of DLP12 lysis genes in Escherichia coli biofilm formation. Microbiology 157, 1640–1650. https://doi.org/10.1099/mic.0.045161-0

Tolar, B., Joseph, L.A., Schroeder, M.N., Stroika, S., Ribot, E.M., Hise, K.B., Gerner-Smidt, P., 2019. An Overview of PulseNet USA Databases. Foodborne Pathog. Dis. 16, 457–462.

https://doi.org/10.1089/fpd.2019.2637

Tonkin-Hill, G., Lees, J.A., Bentley, S.D., Frost, S.D.W., Corander, J., 2018. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. Wellcome Open Res. 3, 93. https://doi.org/10.12688/wellcomeopenres.14694.1

Tran-Dien, A., Le Hello, S., Bouchier, C., Weill, F.-X., 2018. Early transmissible ampicillin resistance in zoonotic Salmonella enterica serotype Typhimurium in the late 1950s: a retrospective, whole-genome sequencing study. Lancet Infect. Dis. 18, 207–214. https://doi.org/10.1016/S1473-3099(17)30705-3

Tsai, K.S., Yang, Y.J., Wang, S.M., Chiou, C.S., Liu, C.C., 2007. Change of serotype pattern of Group D non-typhoidal Salmonella isolated from pediatric patients in southern Taiwan. J. Microbiol. Immunol. Infect. 40, 234–239.

Tsolis, R.M., Xavier, M.N., Santos, R.L., Bäumler, A.J., 2011. How to become a top model: impact of animal experimentation on human Salmonella disease research. Infect. Immun. 79, 1806–1814. https://doi.org/10.1128/IAI.01369-10

Tucker, C.P., Heuzenroeder, M.W., 2004. ST64B is a defective bacteriophage in Salmonella enterica serovar Typhimurium DT64 that encodes a functional immunity region capable of mediating phage-type conversion. Int. J. Med. Microbiol. 294, 59–63. https://doi.org/10.1016/j.ijmm.2003.12.001

Uche, I.V., MacLennan, C.A., Saul, A., 2017. A Systematic Review of the Incidence, Risk Factors and Case Fatality Rates of Invasive Nontyphoidal Salmonella (iNTS) Disease in Africa (1966 to 2014). PLoS Negl. Trop. Dis. 11. https://doi.org/10.1371/journal.pntd.0005118

Uzzau, S., Brown, D.J., Wallis, T., Rubino, S., Leori, G., Bernard, S., Casadesús, J., Platt, D.J., Olsen, J.E., 2000. Host adapted serotypes of Salmonella enterica. Epidemiol. Infect. 125, 229–255. https://doi.org/10.1017/s0950268899004379

van Cappelle, H.G., Veenendaal, D., de Vogel, P.L., 1995. Salmonella panama osteomyelitis in an otherwise healthy patient. A case report. Clin. Orthop. Relat. Res. 235–238.

van Leeuwen, W.J., Voogd, C.E., Guinee, P.A., Manten, A., 1982. Incidence of resistance to ampicillin, chloramphenicol, kanamycin, tetracycline and trimethoprim of Salmonella strains isolated in The Netherlands during 1975-1980. Antonie Van Leeuwenhoek 48, 85–96. https://doi.org/https://doi.org/10.1007/BF00399490

Van Puyvelde, S., Pickard, D., Vandelannoote, K., Heinz, E., Barbé, B., de Block, T., Clare, S., Coomber, E.L., Harcourt, K., Sridhar, S., Lees, E.A., Wheeler, N.E., Klemm, E.J., Kuijpers, L., Mbuyi Kalonji, L., Phoba, M.-F., Falay, D., Ngbonda, D., Lunguya, O., Jacobs, J., Dougan, G., Deborggraeve, S., 2019. An African Salmonella Typhimurium ST313 sublineage with extensive drug-resistance and signatures of host adaptation. Nat. Commun. 10, 4280. https://doi.org/10.1038/s41467-019-11844-z

Varela, G., Ochoa, Aguilar, A., 1953. [Salmonella panama and Escherichia coli 055 in the throat of infants]. Rev. Inst. Salubr. Enferm. Trop. 13, 331–333.

Wagner, P.L., Waldor, M.K., 2002. Bacteriophage Control of Bacterial Virulence. Infect. Immun. 70, 3985 LP – 3993. https://doi.org/10.1128/IAI.70.8.3985-3993.2002

Wahl, A., Battesti, A., Ansaldi, M., 2019. Prophages in Salmonella enterica: a driving force in reshaping the genome and physiology of their bacterial host? Mol. Microbiol. 111, 303–316. https://doi.org/doi:10.1111/mmi.14167

Warrell, D.A., Greenwood, B.M., Davidson, N.M., Ormerod, L.D., Prentice, C.R., 1976. Necrosis, haemorrhage and complement depletion following bites by the spitting cobra (Naja nigricollis). Q. J. Med. 45, 1–22. https://doi.org/https://doi.org/10.1093/oxfordjournals.qjmed.a067448

Warwick, C., Lambiris, A.J., Westwood, D., Steedman, C., 2001. Reptile-related salmonellosis. J. R. Soc. Med. 94, 124–126. https://doi.org/10.1177/014107680109400306

Weigel, L.M., Steward, C.D., Tenover, F.C., 1998. gyrA mutations associated with fluoroquinolone resistance in eight species of Enterobacteriaceae. Antimicrob. Agents Chemother. 42, 2661–2667. https://doi.org/10.1128/AAC.42.10.2661

Wheeler, N.E., Gardner, P.P., Barquist, L., 2018. Machine learning identifies signatures of host adaptation in the bacterial pathogen Salmonella enterica. PLoS Genet. 14, e1007333. https://doi.org/10.1371/journal.pgen.1007333

Whiley, H., Gardner, M.G., Ross, K., 2017. A Review of Salmonella and Squamates (Lizards, Snakes and Amphibians): Implications for Public Health. Pathog. (Basel, Switzerland) 6. https://doi.org/10.3390/pathogens6030038

Wick, R.R., Judd, L.M., Gorrie, C.L., Holt, K.E., 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. 13, e1005595. https://doi.org/10.1371/journal.pcbi.1005595

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

Wilkins, E.G., Roberts, C., 1988. Extraintestinal salmonellosis. Epidemiol. Infect. 100, 361–368. https://doi.org/10.1017/s095026880006711x

Winfield, M.D., Groisman, E.A., 2004. Evolution and Ecology of Salmonella. EcoSal Plus 1. https://doi.org/10.1128/ecosalplus.6.4.6

Wong, V.K., Baker, S., Connor, T.R., Pickard, D., Page, A.J., Dave, J., Murphy, N., Holliman, R., Sefton, A., Millar, M., Dyson, Z.A., Dougan, G., Holt, K.E., Parkhill, J., Feasey, N.A., Kingsley, R.A., Thomson, N.R., Keane, J.A., Weill, F.-X., Le Hello, S., Hawkey, J., Edwards, D.J., Harris, S.R., Cain, A.K., Hadfield, J., Hart, P.J., Thieu, N.T.V., Klemm, E.J., Breiman, R.F., Watson, C.H., Edmunds, W.J., Kariuki, S., Gordon, M.A., Heyderman, R.S., Okoro, C., Jacobs, J., Lunguya, O., Msefula, C., Chabalgoity, J.A., Kama, M., Jenkins, K., Dutta, S., Marks, F., Campos, J., Thompson, C., Obaro, S., MacLennan, C.A., Dolecek, C., Keddy, K.H., Smith, A.M., Parry, C.M., Karkey, A., Dongol, S., Basnyat, B., Arjyal, A., Mulholland, E.K., Campbell, J.I., Dufour, M., Bandaranayake, D., Toleafoa, T.N., Singh, S.P., Hatta, M., Newton, P.N., Dance, D., Davong, V., Onsare, R.S., Isaia, L., Thwaites, G., Wijedoru, L., Crump, J.A., De Pinna, E., Nair, S., Nilles, E.J., Thanh, D.P., Turner, P., Soeng, S., Valcanis, M., Powling, J., Dimovski, K., Hogg, G., Farrar, J., Mather, A.E., Amos, B., Consortium, I.T., 2016. An extended genotyping framework for Salmonella enterica serovar Typhi, the cause of human typhoid. Nat. Commun. 7, 12827. https://doi.org/10.1038/ncomms12827

Wood, D.E., Salzberg, S.L., 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15, R46. https://doi.org/10.1186/gb-2014-15-3-r46

Wood, M.W., Jones, M.A., Watson, P.R., Hedges, S., Wallis, T.S., Galyov, E.E., 1998. Identification of a pathogenicity island required for Salmonella enteropathogenicity. Mol. Microbiol. 29, 883–891. https://doi.org/10.1046/j.1365-2958.1998.00984.x

Worley, J., Meng, J., Allard, M.W., Brown, E.W., Timme, R.E., 2018. Salmonella enterica Phylogeny Based on Whole-Genome Sequencing Reveals Two New Clades and Novel Patterns of Horizontally Acquired Genetic Elements. MBio 9, e02303-18. https://doi.org/10.1128/mBio.02303-18

Yang, J., Barrila, J., Roland, K.L., Kilbourne, J., Ott, C.M., Forsyth, R.J., Nickerson, C.A., 2015. Characterization of the Invasive, Multidrug Resistant Non-typhoidal Salmonella Strain D23580 in a Murine Model of Infection. PLoS Negl. Trop. Dis. 9, e0003839. https://doi.org/10.1371/journal.pntd.0003839

Yao, K., Muruvanda, T., Roberts, R.J., Payne, J., Allard, M.W., Hoffmann, M., 2016. Complete Genome and Methylome Sequences of Salmonella enterica subsp. enterica Serovar Panama (ATCC 7378) and Salmonella enterica subsp. enterica Serovar Sloterdijk (ATCC 15791). Genome Announc. 4. https://doi.org/10.1128/genomeA.00133-16

Yoshida, C.E., Kruczkiewicz, P., Laing, C.R., Lingohr, E.J., Gannon, V.P.J., Nash, J.H.E., Taboada, E.N., 2016. The Salmonella In Silico Typing Resource (SISTR): An Open Web-Accessible Tool for Rapidly Typing and Subtyping Draft Salmonella Genome Assemblies. PLoS One 11, e0147101. https://doi.org/10.1371/journal.pone.0147101

Yoshimura, D., Kajitani, R., Gotoh, Y., Katahira, K., Okuno, M., Ogura, Y., Hayashi, T., Itoh, T., 2019. Evaluation of SNP calling methods for closely related bacterial isolates and a novel high-accuracy pipeline: BactSNP. Microb. genomics 5, e000261. https://doi.org/10.1099/mgen.0.000261

Young, B.C., Earle, S.G., Soeng, S., Sar, P., Kumar, V., Hor, S., Sar, V., Bousfield, R., Sanderson, N.D., Barker, L., Stoesser, N., Emary, K.R.W., Parry, C.M., Nickerson, E.K., Turner, P., Bowden, R., Crook, D.W., Wyllie,

D.H., Day, N.P.J., Wilson, D.J., Moore, C.E., 2019. Panton–Valentine leucocidin is the key determinant of Staphylococcus aureus pyomyositis in a bacterial GWAS. Elife 8, e42486. https://doi.org/10.7554/eLife.42486

Zankari, E., Allesøe, R., Joensen, K.G., Cavaco, L.M., Lund, O., Aarestrup, F.M., 2017. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. J. Antimicrob. Chemother. 72, 2764–2768. https://doi.org/10.1093/jac/dkx217

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 67, 2640–2644. https://doi.org/10.1093/jac/dks261

Zerbino, D.R., 2010. Using the Velvet de novo assembler for short-read sequencing technologies. Curr. Protoc. Bioinforma. Chapter 11, Unit-11.5. https://doi.org/10.1002/0471250953.bi1105s31

Zhang, S., Yin, Y., Jones, M.B., Zhang, Z., Deatherage Kaiser, B.L., Dinsmore, B.A., Fitzgerald, C., Fields, P.I., Deng, X., 2015. Salmonella Serotype Determination Utilizing High-Throughput Genome Sequencing Data. J. Clin. Microbiol. 53, 1685 LP – 1692. https://doi.org/10.1128/JCM.00323-15

Zhou, Z., Alikhan, N.-F., Mohamed, K., Fan, Y., Group, A.S., Achtman, M., 2020. The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. Genome Res. 30, 138–152. https://doi.org/10.1101/gr.251678.119

Zhou, Z., Lundstrom, I., Tran-Dien, A., Duchene, S., Alikhan, N.-F., Sergeant, M.J., Langridge, G., Fotakis, A.K., Nair, S., Stenoien, H.K., Hamre, S.S., Casjens, S., Christophersen, A., Quince, C., Thomson, N.R., Weill, F.-X., Ho, S.Y.W., Gilbert, M.T.P., Achtman, M., 2018. Pan-genome Analysis of Ancient and Modern Salmonella enterica Demonstrates Genomic Stability of the Invasive Para C Lineage for Millennia. Curr. Biol. 28, 2420-2428.e10. https://doi.org/10.1016/j.cub.2018.05.058

Appendix 1

Availability of Supplementary information referred to in this thesis.

Chapter 2

Table S2.1 Reptile associated *Salmonella* strain metadata and control strain metadata
https://figshare.com/s/128e8b3213866db1d78e

Table S2.2 Reptile associated *Salmonella* study contextual reference genomes metadata and accession numbers
https://figshare.com/s/8e36892c2476563777dc

Table S2.3 Reptile associated *Salmonella* QUAST assembly statistics
https://figshare.com/s/a57fd1dcb95e2c9ac0c3

Data S2.1 Reptile associated *Salmonella* carbon source utilisation gene sequences
https://figshare.com/s/874969039f859c62b1c8

Chapter 3

Table S3.1 *Salmonella* Typhimurium metadata and accession numbers
https://figshare.com/s/4c9d4f64d41df3ebb80b

Table S3.2 *Salmonella* Typhimurium quality control statistics
https://figshare.com/s/36ed5e85030bbcd214dc

Table S3.3 *Salmonella* Typhimurium study contextual metadata and accession numbers
https://figshare.com/s/598d16675fc503b71f0f

Table S3.4 *Salmonella* Typhimurium pseudogenes including description, genotype and phenotype
Available overleaf

Table S3.5 *Salmonella* Typhimurium mutations in top predictor genes for invasiveness
https://figshare.com/s/3de71e6a4afecf6cbc76

Data S3.1 *Salmonella* Typhimurium BEAST xml file
https://figshare.com/s/cf0f4bbaefc327505894

Chapter 4

Table S4.1 *Salmonella* Panama metadata and accession numbers
https://figshare.com/s/547b0d8c1b1d12414b55

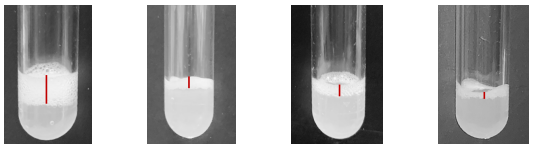
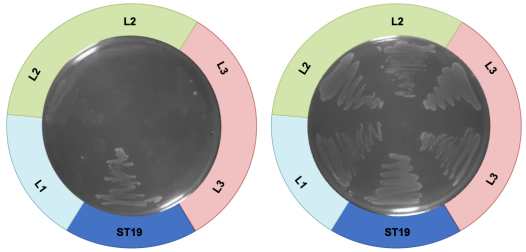
Table S4.2 *Salmonella* Panama GWAS results
https://figshare.com/s/c536c03e527aaab12acd

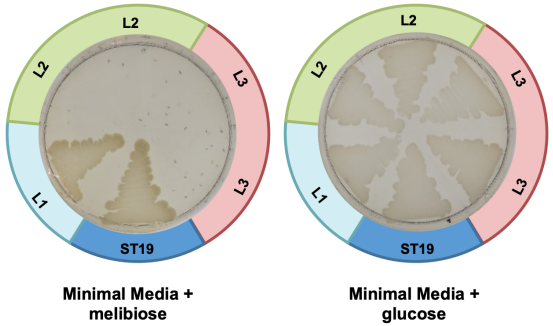
Table S4.3 *Salmonella* Panama functional annotation of prophage
https://figshare.com/s/812e6b96bb5105f4a18d

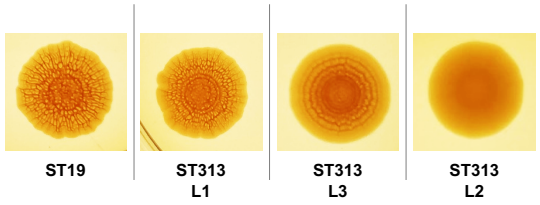
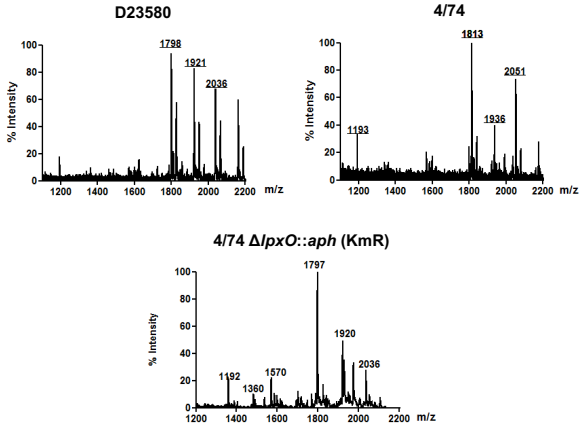
Data S4.1 *Salmonella* Panama BEAST xml file
https://figshare.com/s/44debf3701ada5d90d20

Data S4.2 *Salmonella* Panama GWAS genes
https://figshare.com/s/59f81f9ce500159c3c39

Data S3.4 *Salmonella* Typhimurium pseudogenes including description, genotype and phenotype

Gene Name	Gene Description	Genotypic Changes	Further evidence	ST19*	ST313 L1*	ST313 L3*	ST313 L2*
***ratB* (STM2514)**	Secreted outer membrane protein associated with gut persistence in murine models.	ST313 L1: Q657*, Q1067H, T885I ST313 L2: Q657*, Q1067H, R66C ST313 L3: Q657*, Q1067H, R66C, G820S Abundance of SNPs occur sporadically throughout the phylogeny.	Although a causal relationship is unproven, inactivation of *ratB* suggested to have reduced the enteric potential of ST313, resulting in more systemic infections (Okoro et al., 2015).				
***katE* (STM1318)**	Stationary phase catalase involved in protecting high density bacterial communities in the environment from oxidative stress.	ST313 L1: E117G ST313 L2: E117G ST313 L3: E117G Synonymous SNPs are additionally present in ST313 L1 (A→G at base 1452) and ST313 L3 (A→G at base 1908).	Reduction in catalase activity across all ST313 (Ashton et al., 2017; Singletary et al., 2016). ST313 L3 had lower catalase activity than ST19 (this study).  **ST19 Str. 474** **ST313 L1 Str. A130** **ST313 L3 Str. BKQ ZM9** **ST313 L2 Str. D23580** *bubbles represent catalase activity				
***ttdA* (STM3355)**	L(+)-tartrate dehydratase involved in tartaric acid utilisation.	ST313 L1: Q68* ST313 L2: Q68* ST313 L3: Q68* Additional mutations sporadically present across the phylogeny.	L-tartaric acid and dihydroxyacetone cannot be used as sole carbon sources by ST313. Results were originally generated with Biolog phenotype microarrays (Okoro et al., 2015). ST313 L3 was also unable to grow on tartaric acid as a sole carbon source (this study).  				

***melR* (STM4297)**	Melibiose operon response regulator involved in melibiose utilisation.	ST313 L2: F311L ST313 L3: F311L An additional synonymous SNP was also identified at base position 462 in ST313 L3.	ST313 L2 cannot grow on melibiose as a sole carbon source (Yang et al., 2015). Assessment of the alpha-galactosidase activity of ST19 and ST313 wild-types and mutant strains showed that *melR* controlled the melibiose utilisation system (Canals et al., 2019). ST313 L3 was unable to grow on melibiose as a sole carbon source (this study).  				
***flhA* (STM1913)**	Transmembrane biosynthesis protein involved in flagella protein export.	ST313 L2:A166T ST313 L3:A166T	The A166T SNP caused reduced motility in ST313 L2, demonstrated by comparing D23580 *flhA*^{474} mutant to wild type (Canals et al., 2019).				
***pipD* (STM1094)**	Pathogenicity island-encoded protein D.	ST313 L2: 283bp deletion ST313 L3: 283 bp deletion plus additional 109 bp deletion.	Although a causal relationship is unproven, the *pipD* gene contributes to macrophage persistence in murine models (Lawley et al., 2006; Okoro et al., 2015). Of further interest, the *Salmonella* pathogenicity island SPI-5 gene *sopB* encodes a SPI-1 effector and has an associated chaperone *pipC.* In ST19 strains, these genes are upregulated in the gastrointestinal tract, which facilitates invasion into epithelial cells. However, in ST313 strains *sopB* and *pipC* are also upregulated in SPI-2 media and macrophages (Canals et al., 2019). It is possible that this regulatory change is due to the loss of *orfX,* STM1093 and the 3' end of *pipD.*				

bcsG **(STM3624)**	Cellulose biosynthetic enzyme involved in biofilm formation.	ST313 L2: W247*	The RDAR phenotype of ST313 L2 is negative due to a mutation in *bcsG* (Singletary et al., 2016). The RDAR phenotype of ST313 L3 has an intermediate RDAR phenotype, however the genetic basis is unknown (this study).  *RDAR colony morphology at 25°C				
lpxO **(STM4286)**	Putative membrane-bound beta-hydroxylase.	ST313 L2: E198* An additional, synonymous SNP was also present across all ST313 lineages (G→T at base 411).	In D23580, *lpxO* is a pseudogene due to the presence of one SNP introducing an early STOP codon (Kingsley et al., 2009). By mass spectrometry, it was confirmed that the lipid A of D23580 was not modified by LpxO during growth in InSPI2 medium, while the lipid A of 4/74 contained this modification when grown in this medium. To confirm that LpxO-mediated modification of lipid A, it was established that the mass spectrum of the lipid A from a 4/74 Δ*lpxO*::*aph* (KmR) mutant grown in InSPI2 was similar to the spectrum of D23580 grown in the same medium (this study).  				

***sseI* (STM1051)**	Type III secretion system effector protein (SPI-2) involved in host cell dissemination.	In ST313 L2, the *sseI* gene contains an IS*26* transposase insertion. Generally, ST19, ST313 L1 and ST313 L3 strains carry a functional *sseI* gene. However, there are several ST313 L1 genomes which contain the IS*26* transposase. *sseI* mutations are abundant and sporadic throughout all ST313 lineages. Specific non-synonymous mutations commonly occur in at least 10 amino acid positions; V13N, T47A, V69I, V70N, D81A, N86S, L91Q, D114Y, V115A and R266G.	Studies in mice have demonstrated that an accumulation of SNPs within *sseI* in ST313 L2 leads to an increased ability of ST313 L2 bacteria to disseminate rapidly from the gut to the draining lymph nodes in a murine model (Carden et al., 2017).				
***macB* (STM0942)**	Putative ABC superfamily transport protein required to resist peroxide mediated killing	ST313 L1: W262* ST313 L2 sub-lineage: Deletion of A and C at nucleotide position 444 and 445.	MacAB is involved in oxidative stress resistance (Bogomolnaya et al., 2013) and mediates resistance to macrolide antibiotics (Nishino et al., 2006). Recently, ST313 L2-associated variants of the MacAB-TolC tripartite efflux pump have been shown to affect replication in macrophages, and influence fitness during colonisation of the murine gastrointestinal tract (Honeycutt et al., 2020).				

Note: Phenotypic testing was conducted to determine the relevance of functional gene loss by B. Perez-Sepulveda, R. Canals, N. Wenner, X. Zhu, H. Webster (University of Liverpool, UK), J. Bengoechea and A. Dumigan (Queens University Belfast, UK).

* Note that predicted functionality (see methods) is depicted as a colour strip for each gene and is based on whole genome-based predictions of SNPs likely to play a functional role. Specifically, blue indicates functional and pink indicates non-functional.

** Mutations refer to position in individual genes taken from D23580, non-synonymous mutations are listed as amino acid changes including amino acid positions and synonymous changes are listed as nucleotide changes with nucleotide positions.